
THE UNIVERSITY
*of* EDINBURGH

# THE ROLE OF LSH IN THE ESTABLISHMENT OF EPIGENETIC GENE SILENCING

Natalia Torrea Muguerza

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2018


THE UNIVERSITY *of* EDINBURGH

## Declaration

I declare that this thesis was composed by myself. The research in this thesis is my own, unless otherwise stated and has not been submitted for any other degree or professional qualification.

Natalia Torrea (July 2018)

# Table of contents

## Acknowledgements

First of all, I would like to thank Irina for giving me the opportunity to join her lab and later on carry out my PhD research under her supervision. I would like to thank her for the advice throughout my project, for interesting discussions, for helpful ideas and for her encouragement to keep my motivation. Also, thank you for reading the draft of my thesis and for the constructive feedback provided. I am really grateful for the professional and emotional support received during this time.

I would also like to thank The Darwin Trust of Edinburgh, who funded my PhD during the last four years.

Additionally, I would like to thank my current supervisors Patrick and Philipp for their support and feedback on my thesis project, especially during the final months of my PhD. Thank you to Sara for numerous discussions and the members of my thesis committee, Liz and Alastair, for their feedback and comments.

Next, I would like to acknowledge current and past members of the Stancheva lab. Thank you Chao, Tuo and Ausma, for providing me with the material, abilities and background knowledge required for my thesis project. I would like to thank Jasmin and Eboni, former students in the lab, for their help carrying out some of my experiments. A special thank you to Burak, Simon, Christian, Ilaria, Dani and Elana, for their support in the lab and outside the lab. We had so much fun together and enjoyed needed breaks during the hard work. Thank you for contributing to a great professional and personal time in Edinburgh.

Thank you to my friends: Ilaria, Pablo, Elana, Natalia, Tülin, Covi and Mayte, who have been like a family in Edinburgh. I would also like to thank Alba for her help in the important moments and for all the hard but fun work developing our own outreach activity. Thank you to Sarah and Maria for all the exciting public engagement events I was able to help with.

I would like to thank Dani for being there in the last years, for helping with images for my thesis and for much needed time away from the PhD.

Finally I would like to thank my family and I need to thank them in Spanish. Gracias a mis padres por su ayuda en todos los aspectos de mi vida y por saber que siempre puedo contar con ellos. Gracias por el apoyo en mis estudios y etapas en el extranjero que me llevaron a empezar la tesis en Edimburgo. Gracias a mi hermana por estar siempre ahí.

## Abstract

DNA methylation is essential for mammalian development and transcriptional repression of genes and retrotransposons during embryo development and in somatic cells. The patterns of DNA methylation are established by *de novo* DNA methyltransferases, which are regulated by developmental signalling and require access to chromatin. Besides DNA methyltransferases, other proteins have recently been implicated in DNA methylation, such as the ATP-dependent chromatin remodeler LSH. The absence of LSH in mouse embryos leads to defects in DNA methylation and development. In relation to this, mutations in LSH have been found to cause Immunodeficiency–Centromeric instability–Facial anomalies (ICF) syndrome. This syndrome is characterized by centromeric instability and CpG hypomethylation of centromeric satellite repeats, and is most often caused by mutations in the catalytic domain of the DNA methyltransferase DNMT3B.

LSH is essential for developmentally programmed *de novo* DNA methylation of large chromosomal domains including promoters of protein coding genes and repetitive sequences. Importantly, fibroblasts derived from chromatin remodeling ATPase LSH-null mouse embryos, which lack DNA methylation at transposons and specific gene promoters, are capable of re-establishing normal patterns of DNA methylation and transcriptional silencing of misregulated genes upon re-expression of LSH. The ATP hydrolysis by LSH is essential for its function in gene silencing and *de novo* DNA methylation. However, the molecular mechanisms of LSH-dependent gene silencing and *de novo* DNA methylation are yet unclear. Here we use an inducible system that enables controlled expression of LSH in *Lsh*-null mouse embryonic fibroblasts (MEFs) to follow chromatin dynamics, transcriptional silencing and establishment of *de novo* DNA methylation. This conditionally reversible *Lsh* knockout cellular system allowed us to study the order of events occurring immediately after LSH restoration in MEF cell lines in order to elucidate the molecular mechanism of LSH-dependent gene silencing. We have demonstrated that LSH upon its restoration localises to the promoters of LSH-dependent loci leading to a mild decrease in the occupancy of H3, which reinforces the previously shown role of LSH as a chromatin remodeler. Simultaneously, there is removal of acetyl groups from H3 tails when LSH is bound to these target regions, which might be facilitated by the interaction of HDACs with LSH. The removal of H3Ac marks is followed by deposition of H3K9me2 by G9a/GLP histone methylases at the same time point when misregulated genes are silenced. This suggests that LSH creates a suitable substrate for G9a/GLP promoting gene silencing. Surprisingly, transcriptional repression occurs without acquisition of DNA methylation at the promoters of these loci. This order of events implies that LSH plays a role as a chromatin remodeler leading to changes in chromatin structure and modifications that facilitate epigenetic gene silencing without DNA methylation in the initial

period when LSH is restored in MEF cell lines. Furthermore, deposition of H3K9me2 by the G9a/GLP complex is critical for silencing of specific genes, but not for repetitive elements such as IAPs. The histone modification H3K27me3 seems to play a transitory role in the silencing of IAP retrotransposons in the absence of G9a/GLP activity.

In conclusion, this work has demonstrated that changes in chromatin modifications leading to a transcriptionally repressive chromatin state can be established in somatic cells by the chromatin remodeler LSH without acquisition of DNA methylation. This suggests that the primary role of LSH is to promote changes in chromatin structure and modifications that lead to gene silencing and not DNA methylation, which most likely occurs as a consequence of transcriptional silencing.

# Chapter 1 - Introduction

## 1.1 DNA packaging into chromatin

Mammalian diploid cells contain approximately 6 billion base pairs of DNA, corresponding to about 2 meters that need to be packaged to fit into the nucleus. The basic level of DNA compaction consists of 147 bp of DNA double helix wrapped around a histone protein octamer in 1.65 left-handed turns to form a nucleosome unit with DNA compacted into a superhelix. The nucleosome unit represents the basic structure of chromatin (Kornberg, 1974; Luger, Mäder, Richmond, Sargent, & Richmond, 1997; C. L. F. Woodcock, Safer, & Stanchfield, 1976). The histone octamer consists of one H3/H4 tetramer and two H2A/H2B dimers. Histone N-terminal and C-terminal tails extend from the nucleosome core particle and this facilitates their modification by a series of chromatin regulators (Luger et al., 1997; Luger & Richmond, 1998). All histone can be post-translationally modified (PTM) in a dynamic and reversible manner and this contributes to the regulation of chromatin function (Bannister & Kouzarides, 2011). The nucleosome core particles are interconnected by linker DNA, which varies in length depending on species and tissue. This linker DNA is associated with the linker histone H1 (Bednar et al., 1998). The 10nm chromatin fibre consists of nucleosomal arrays which are regularly spaced nucleosomes bound by linker histone H1 and other nucleosome binding proteins (Luger & Hansen, 2005). Interactions of linker DNA and linker histone H1 were found to result in a unique structural motif that directs chromatin folding and compaction facilitating the folding of chromatin into a 30 nm fibre (Bednar et al., 1998; C. L. Woodcock, 1994). Depletion of histone H1 in embryonic stem cells has been shown to cause a decrease in the spacing of nucleosomes and a reduction in the compaction of chromatin. Therefore H1 seems to be essential for higher order chromatin folding and chromosome structure (Y. Fan et al., 2005). Some proteins which are members from the high-mobility group protein superfamily, weaken the binding of H1 to nucleosomes and cause a decrease in the degree of compaction of the chromatin fibre. Therefore these proteins tend to facilitate the access of regulatory factors to their chromatin targets (Catez et al., 2004), which suggests that a network of structural proteins that interact constantly, modulates the nucleosome accessibility and the local structure of the chromatin fibre. However, it is not clear whether the 30 nm fibre exists in higher eukaryotes since there is no strong *in vivo* evidence for it, although it can be assembled *in vitro* (only on repetitive and very regular DNA sequences) and extracted from nuclei. This suggests that the 30 nm fibre could be a distinct secondary higher-order chromatin structure, while an unfolded 30 nm fibre intermediate might be the preferred conformation of chromatin *in vivo* (Tremethick, 2007). It has been shown that the nucleosomal arrays, the 10 nm fibre, can be condensed into multiple higher order structures, which are defined as secondary and

tertiary chromatin structures (Luger & Hansen, 2005). The secondary structures refer to the 30 nm chromatin fibre, which is formed by interactions of nucleosomes and regulatory proteins. The interaction between the N-terminal tail domain of histone H4 and the surface of histone H2A from a neighbouring nucleosome is needed to form this secondary chromatin structures (Dorigo, Schalch, Bystricky, & Richmond, 2003; J. Y. Fan, Rangasamy, Luger, & Tremethick, 2004). The tertiary structures, which are formed by interactions between secondary structures, include long distance contacts involving enhancers and promoters or looped chromatin domains (Woodcock and Dimitrov 2001). Many nucleosome-binding proteins have the capacity to assemble specific secondary and tertiary chromatin structures. These chromatin structures, formed by interactions with non-histone proteins, differ from the structures formed during intrinsic fibre folding (Luger & Hansen, 2005). For example, heterochromatin protein 1 α (HP1α) has been shown to bind to folded nucleosomal arrays inducing local rearrangement of secondary chromatin structures (J. Y. Fan et al., 2004). Besides the canonical histones H2A, H2B, H3, and H4 deposited immediately after DNA replication to package the newly generated genome, there are histones variants that can influence the properties and dynamics of the nucleosome. These histone variants play an important role during transcription influencing gene regulation (Weber & Henikoff, 2014). Furthermore, it has been shown that HP1α binds preferentially to condensed higher order chromatin structures and cooperates with the histone variant H2A.Z to generate highly compacted secondary structures of chromatin (J. Y. Fan et al., 2004).

Compaction of chromatin into higher order structures is thought to impair the access of regulatory factors to DNA, but not all the genome is condensed to the same level. This allows different regions of the genome to be regulated in different cells and at different stages during development (Trojer & Reinberg, 2007). Based on a classic definition there are two distinct states of chromatin configuration, active chromatin or euchromatin and silent chromatin or heterochromatin. Euchromatin usually contains single copy genes and genome regions that are transcribed, while heterochromatin contains repetitive sequences and regulates various functions including gene silencing, normal centromere function and nuclear organization (Bird 2002). Heterochromatin can be divided into two different types - constitutive and facultative heterochromatin. Constitutive heterochromatin includes repetitive and non-coding sequences such as centromeric and telomeric DNA. On the other hand, facultative heterochromatin includes condensed and transcriptionally silent chromatin regions that retain the capacity to interconvert with euchromatin depending on the specific circumstances, such as developmental stage or nuclear localization (Trojer & Reinberg, 2007). Besides euchromatin and heterochromatin, there is now evidence for diversity of genome segmentations into different chromatin states (Filion, Bemmel, Braunschweig, & Talhout, 2010). Also, targeting of

DNA-binding factors to specific regions is influenced by the different structures of chromatin (Filion, van Bemmel et al. 2010).

In summary, the dynamics of nucleosomal compaction allow to access the DNA when needed while DNA is simultaneously packaged into highly condensed chromosomal structures. The different structures of chromatin are dynamic and allow transitions between different compaction states. Numerous proteins and protein complexes target nucleosomes in chromatin fibres to disassemble, reassemble and remodel nucleosome structures (Luger & Hansen, 2005). The effects of proteins, protein complexes and modifications on chromatin structure are not independent processes. There is an interplay between them. One example is the linker histone H1 the binding of which has been shown not only to reduce the mobility of H3 tails but also inhibit modifications of H3. Likewise, modifications of histones such as phosphorylation and acetylation, which weaken the contact between H3 tail and linker DNA, facilitate H3 tail dynamics and modifications (Stützer et al., 2016). Finally, the chromatin dynamics are also regulated by DNA methylation and ATP-dependent enzymes which remodel chromatin in addition to the post-translational modifications of histone proteins mentioned earlier (Ho & Crabtree, 2010). These have implications in different processes such as gene regulation, transcription and DNA repair (Luger, 2003; Luger & Hansen, 2005).

## 1.2 Epigenetic modifications and gene regulation

Waddington referred to epigenetics as the study of how genotypes give rise to different phenotypes during development (Waddington, 1957). This was followed by the definition of epigenetics as the study of heritable changes in the gene function, which cannot be explained by the changes of the genetic sequences (Russo V.E.A., Martienssen R.A., & Riggs A.D., 1996). There are different layers of epigenetic regulation including DNA methylation, post-translational modifications of amino terminal tails of histone proteins and exchange of canonical histones by histone variants (A. Bird, 2007). The inheritance of epigenetic profiles is necessary to maintain the gene expression characteristic of different cell types during cell division (M. Xu & Zhu, 2010), however not all epigenetic marks are transmissible between generations. Examples of this are transient phosphorylation of the histone variant H2A.X after a double-strand break and histone modifications associated with transcription (A. Bird, 2007). Adrian Bird refined the classical definition, referring to epigenetics as “the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states”.

Gene activation is regulated by transcription factors and gradients of signalling molecules and epigenetic modifications can contribute to stabilise the expression patterns. In addition, gene silencing is also controlled by epigenetic mechanisms involving post-synthetic

modifications of DNA, such as DNA methylation, and modifications of amino-terminal tails of histone proteins (Klose & Bird, 2006; Leeb & Wutz, 2012). Some proteins bind specifically to methylated DNA and promote changes in chromatin acting as a link between DNA methylation and chromatin remodeling and modification (Klose and Bird 2006). DNA methylation, histone modification and chromatin remodelling are not independent processes, they are linked through different molecular mechanisms. This has important consequences for developmental and differentiation processes causing alterations in appropriate gene expression patterns that can lead to diseases and cancer through silencing of cancer-relevant genes (P. a Jones & Baylin, 2007). An example of the interaction between DNA methylation and histone modifications are proteins that contain a methyl CpG binding domain (MBD), which can recruit chromatin modifiers, such as histone deacetylases, that condense chromatin and repress transcription (Wade & Wolffe, 2001). Another layer of epigenetic regulation is provided by the exchange of histone variants that can occur independently of DNA replication. Non-canonical histone variants play an important role during transcription and might have a role perpetuating active chromatin states (Weber & Henikoff, 2014). The inheritance of all epigenetic marks has been questioned, but some epigenetic profiles, particularly the repressive ones, are maintained through DNA replication. Histone post-translational modifications in constitutive pericentromeric heterochromatin are maintained during chromatin replication (Cedar & Almouzni, 2016). These epigenetic marks not only help maintaining gene expression profiles but also maintaining chromatin structure such as centromere and telomere through interactions between specific histone modifications and chromatin binding proteins (M. Xu & Zhu, 2010). The next sections will describe in detail the main epigenetic marks that will be studied in this work, DNA methylation and post-translational histone modifications (Figure 1.1).

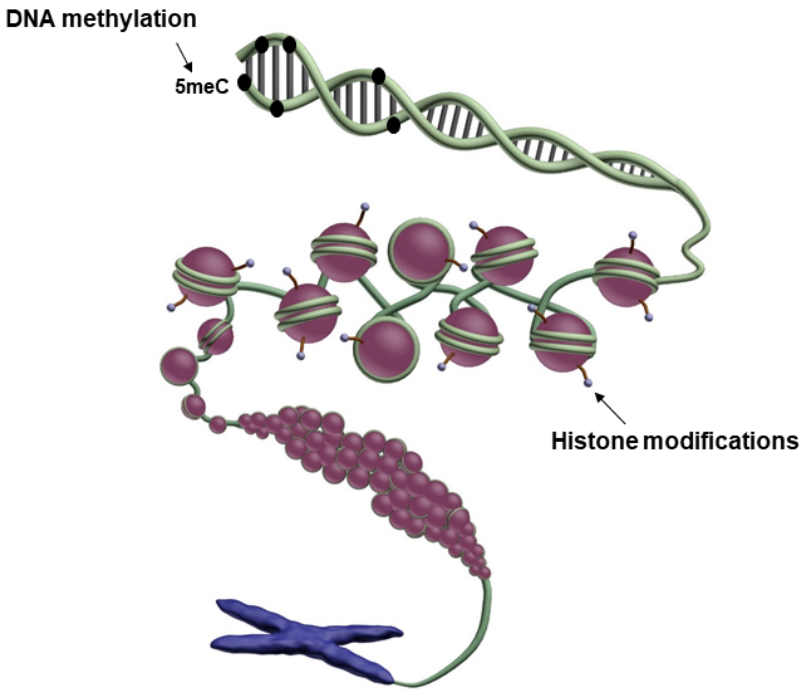


**Figure 1.1. Epigenetic modifications on DNA compacted into chromatin.**

DNA double helix wrapped around core histones into nucleosome units, that are further compacted into higher order chromatin structures. DNA methylation at CpG dinucleotides (5meC) indicated as black circles on the DNA double helix. Post-translational histone modifications indicated as circles on protruding histone protein N-terminal tails.

## 1.3 DNA methylation

Multiple epigenetic mechanisms affect the structure of chromatin and lead to activation or repression of some genes. DNA methylation is one of the mechanisms that play a role in chromatin structure (Theresa M Geiman & Muegge, 2010; Hansen, Ghosh, & Woodcock, 2010). In animals, DNA methylation refers to the addition of a methyl group to the 5-position carbon within the cytosine pyrimidine ring. During the methylation process, DNA methyltransferases (DNMTs) use S-adenosylmethionine (SAM) as a methyl group donor to transfer the methyl group from SAM to the cytosine residue which becomes a 5-methyl cytosine (5meC). This process occurs mostly in the context of a cytosine-guanine dinucleotide (CpG) in the mammalian genome (Yoder, Soman, Verdine, & Bestor, 1997). Methylation can also occur in non-CpG dinucleotide context at cytosines in the context of CH, CHG and CHH, where H is a non-guanine nucleotide. Methylation of cytosine at these other contexts is widespread in plants (Cokus et al., 2008), and has recently been observed in mammals (Lister

et al., 2009). However, the function of non-CpG methylation is currently unknown in mammals (P. A. Jones, 2012) although, like CpG methylation, non-CpG methylation has been suggested to have a function in gene silencing (Malone et al., 2001). During DNA replication, the methylation mark is passed to the newly synthesised DNA strands due to the symmetric nature of the CpG methylation in the genome (A. Bird, 2002). In somatic cells, only 1% of all DNA bases is a methyl-cytosine but about 70-80% of the CpGs in the genome are methylated (Ehrlich et al., 1982). An exception to this highly methylated CpGs in the genome are the CpG islands (CGI). CpG islands are regions high in CpG dinucleotide content which are usually unmethylated (A. P. Bird, 1986). The length of CpG islands varies between hundreds of base pairs and four kilo-base pairs. Most CpG islands are positioned in the proximal, 5' end, promoter region of the genes in the mammalian genome and are normally unmethylated (A. Bird, 2002; A. Bird, Taggart, Frommer, Miller, & Macleod, 1985). Approximately 60% of annotated mammalian genes have CpG islands found near the transcription start sites (TSS). Most of them are unmethylated during all stages of development in different tissues types (Antequera & Bird, 1993; A. Bird et al., 1985). The high content of CpG dinucleotides in CpG islands compared to the rest of the genome is due to the conversion of methylated cytosines to thymidines through spontaneous deamination. This explains the low frequency of CpGs found in the genome, in opposition to CpG islands which are normally unmethylated (Antequera & Bird, 1999; Smallwood et al., 2012).

### 1.3.1 Distribution of DNA methylation and function in gene regulation

The distribution of DNA methylation is related to the different functions associated with this epigenetic mark (Figure 1.2). The association of CpG islands with gene promoters suggests that their methylation state has a role in the regulation of gene transcription (A. Bird, 2002). As mentioned earlier, most CpG islands are unmethylated, but a small proportion become methylated during early mammalian development and in disease such as cancer, which leads to long term stable gene silencing in somatic cells (Antequera & Bird, 1999). This long term stability of gene silencing due to methylated CpG islands is also associated with imprinted genes and genes located on the inactive X chromosome. On the other hand, transcriptionally active genes that contain unmethylated CGIs at their promoters are characterized by nucleosome-depleted regions (NDRs) at the TSS. The nucleosomes flanking NDRs often contain the histone variant H2A.Z and are marked with active histone modifications such as the trimethylation of lysine 4 from histone H3 (H3K4me3) and acetylation of histones H3 and H4 (Kelly et al., 2010). The histone modification H3K4me3 impairs the establishment of DNA methylation since DNMTs cannot interact via their ADD domain with methylated H3K4, showing again a link between methylation and histone

modifications (Otani et al., 2009; Sormani, Haerter, Lövkvist, & Sneppen, 2016). The idea of the establishment of DNA methylation at gene promoters that precedes and directs gene silencing has been controversial, but this timing of events might not be the most general mechanism. It seems that generally gene silencing precedes *de novo* DNA methylation at promoters (P. A. Jones, 2012). One example of this timing is the silencing of the *Oct-3/4* gene by the repressive histone modification H3K9me2 and subsequent establishment of DNA methylation during early embryogenesis (Athanasiadou et al., 2010; Feldman et al., 2006). During development, genes containing CpG island promoters can be repressed by Polycomb proteins, such as the PRC2 complex that deposits H3K27me3, a histone modification associated with repressed genes (Curie & Paris, 2011). There is now evidence demonstrating that approximately half of the CpG islands in mammalian genomes are not associated with annotated promoters and they are found in intragenic or intergenic regions. These CpG islands are referred to as orphan CGIs. Orphan CGIs show transcriptional activity although they are more likely to become methylated during development than CGIs at promoters. This suggests a possible functional role of orphan CGIs during development (Illingworth et al., 2010).

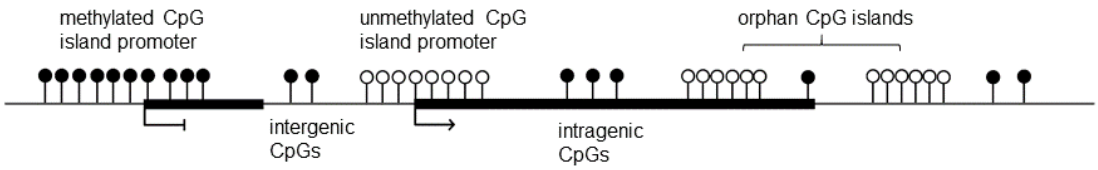


**Figure 1.2. Distribution of DNA methylation at CpG dinucleotides.**

Methylated and unmethylated CpG dinucleotides are indicated as black and white lollipops, respectively. Most of the genome is methylated at CpG dinucleotides (intergenic and intragenic CpG dinucleotides). Exception to this are CpG islands (CpG rich regions) which are usually located at promoter regions. CpG islands are usually unmethylated, unless they are involved in developmental regulation of gene activity. Orphan CpG islands are not associated with annotated promoters and they are found in intragenic or intergenic regions.

Two models have been proposed to explain how DNA methylation is associated with gene silencing. DNA methylation can repress transcription by blocking transcriptional activators from binding to methylated CpGs and also promoting the binding of proteins with a methyl-CpG binding domain (MBD) that can recruit corepressors to silence gene expression. MBD1 can interact with the histone methyltransferase enzyme SETDB1, creating a link between

recognition of DNA methylation and modification of the chromatin state by methylation of histone H3 lysine 9 (Sarraf & Stancheva, 2004). DNA methyltransferases show another layer of regulation, since they interfere with gene expression through interaction with proteins such as histone deacetylases that create a repressive chromatin state (Klose & Bird, 2006). It seems that gene silencing is not caused by a single mechanism such as DNA methylation, but requires multiple different processes for efficient repression including: interference with binding of transcription factors, recruitment of histone deacetylases via repressor proteins, methyl-CpG binding proteins or DNA methyltransferases, and non-coding RNAs (Klose and Bird 2006). These mechanism indicate a crosstalk between different epigenetic marks to stabilise expression states (Sormani et al., 2016).

Most gene bodies are poor in CpG dinucleotide content but CpGs found in gene bodies are methylated. Methylation of gene bodies is associated with silencing of repetitive and transposable elements, such as intracisternal A particle (IAP) elements (Yoder, Walsh, & Bestor, 1997). It has also been shown that intragenic DNA methylation can prevent spurious transcription through a crosstalk between the histone modification H3K36me3 and DNA methyltransferase DNMT3B. Transcribed genes contained H3K36me3 at their gene bodies, which is associated with transcription elongation and can also recruit DNMT3B (Neri et al., 2017). When DNA methylation is established in the gene body, it does not block transcription and it might even stimulate transcription elongation (Wen et al., 2014). Another characteristic of methylation at gene bodies is that exons are highly methylated compared to introns, which correlates with higher nucleosome occupancy at exons (Chodavarapu et al., 2010). The change in the level of DNA methylation between exons and introns occurs at the boundary suggesting a role of DNA methylation in the regulation of splicing (L. Laurent et al., 2010).

Enhancers are CpG poor regions that control gene expression and have variable levels of DNA methylation with an average of 30% 5meC. These regions are referred to as low-methylated regions and this suggests that enhancers might be in a dynamic state since they can be methylated or unmethylated at different times. Transcription factors and proteins that affect DNA methylation levels, such as TET enzymes, have been found at enhancers, but the relation between methylation of enhancers and the regulation of their function is not clear yet (Stadler et al., 2011; Yin et al., 2017). The TET enzymes could be recruited through its CXXC domain to unmethylated CpGs and thus reinforce the unmethylated state (Sormani et al., 2016).

In summary, promoter hypomethylation is associated with potential for active transcription and higher levels of gene body methylation are correlated with active transcription (L. Laurent et al., 2010). DNA methylation is critical in different processes, for example genomic imprinting causing allele-specific expression of some genes (Reik & Walter, 2001), X inactivation

ensuring expression of genes from only one X chromosome in female mammals (Senner and Brockdorff 2009) and silencing of retrotransposons in the genome contributing to genome stability (Huang et al., 2004). Aberrant hypermethylation of CpG islands at gene promoters is important in cancer initiation and progression, since it leads to inappropriate silencing of genes, such as tumour suppressor genes or *BRCA1* gene in breast cancer (P. A. Jones & Baylin, 2002; P. a Jones & Baylin, 2007). Early differentiation is marked by important changes in DNA methylation and genome structure. However, it is still unclear how DNA methylation is stablished at specific sites in the genome and how DNA methylation contributes beyond silencing gene expression to affect chromatin structure, modifications and nuclear architecture (Geiman and Muegge 2010).

### 1.3.2 DNA methyltransferases

DNA methylation patterns are established during gametogenesis, embryo development, and cell differentiation by enzymes of the DNA cytosine methyltransferase family, which includes five DNA methyltransferases (DNMTs) in mammals. The maintenance DNA methyltransferase DNMT1 and the *de novo* methyltransferases DNMT3A and DNMT3B are referred to as the canonical DNMT enzymes since they catalyse the addition of methylation marks to CpGs in the genomic DNA (Okano, Bell et al. 1999, Bird 2002). The functions of these enzymes are not limited to the establishment and maintenance of DNA methylation patterns, they also have an important role in epigenetic gene regulation (Lyko, 2017). The non-canonical DNMT enzymes, DNMT2 and DNMT3L (DNMT3-like), have sequence conservation with the canonical DNMT3 enzymes.

#### Domain architecture of DNMT enzymes

DNMT enzymes share a similar domain structure with an N-terminal regulatory domain and a C-terminal catalytic domain, except for DNMT2 that only contains the catalytic domain and DNMT3L that is catalytically inactive (Figure 1.3). The catalytic domain contains six conserved motifs involved in SAM cofactor binding and catalytic activity (Bestor, 2000). The N-terminal regulatory domains are different between DNMT1 and DNMT3 enzymes and this differences could play an important role in their specific activity in the genome since the N-terminal region is responsible for protein-protein interactions and targeting of these enzymes (Ambrosi, Manzo, & Baubec, 2017).

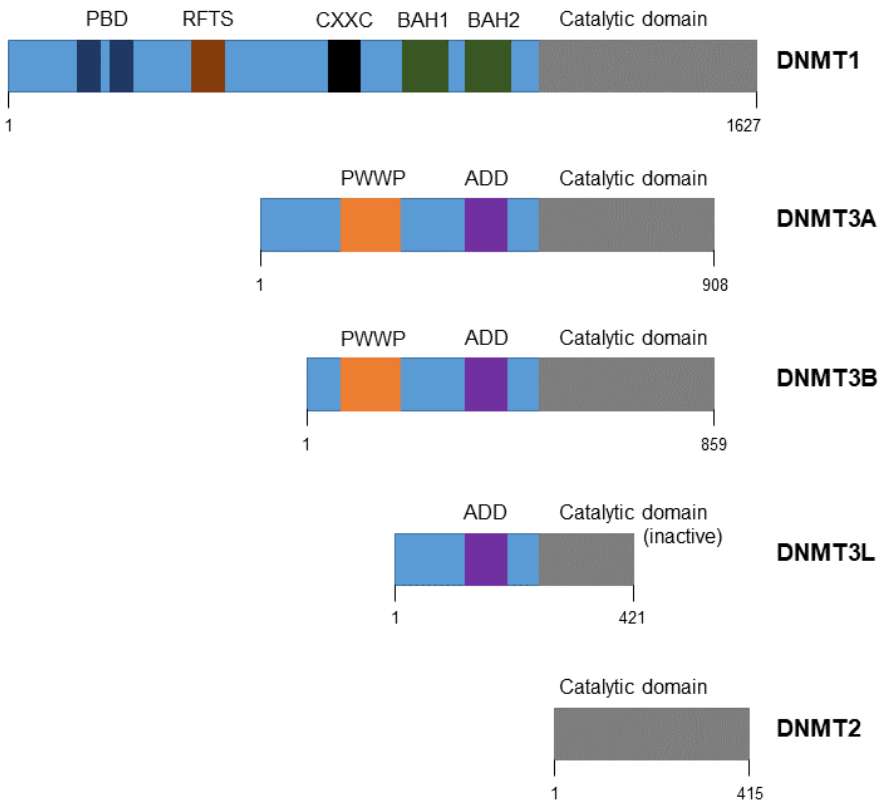


**Figure 1.3. Domain architecture of murine DNMT enzymes.**

Domain annotation of maintenance DNA methyltransferase DNMT1, *de novo* DNA methyltransferases DNMT3A and DNMT3B, and the non-canonical enzymes DNMT3L and DNMT2. PBD: includes the DMAP1 and PCNA binding domains, RFST: replication foci targeting sequence, CXXC: DNA binding domain, BAH: bromo-adjacent homology domains, PWWP: required for chromatin targeting, ADD (ATRX-DNMT3-DNMT3L): important for protein interactions.

The N-terminal domain of DNMT1 contains several conserved subdomains that are important for molecular interactions, such as the protein binding domain (PBD) that is able to interact with different proteins, RFTS domain, CXXC domain and BAH domains. The PBD domain contains a DMAP1 binding domain that is required for DNMT1 to establish a repressive transcription complex during DNA replication in addition to its role maintaining DNA methylation. DNMT1 interacts through this domain with the transcriptional repressor DMAP1 (DNMT1-associated protein 1) and with the histone deacetylase HDAC2 (Rountree, Bachman, & Baylin, 2000). The PBD domain also contains a PCNA binding domain. The RFTS domain, replication foci targeting sequence, is necessary for targeting DNMT1 to the replication foci. This allows the methylation of newly synthesized DNA in a replication-dependent manner and

maintains the specific pattern of methylation (Hervouet, Peixoto, Delage-mourroux, Boyer-guittaut, & Cartron, 2018; Leonhardt, Page, Weier, & Bestor, 1992). The CXXC is a conserved zinc-finger domain that binds unmethylated CpG and the function of the BAH domains, bromo-adjacent homology domains, still needs to be elucidated (Lyko, 2017).

The N-terminal region of DNMT3A and DNMT3B enzymes also contains conserved domains, including a PWWP domain and an ADD domain. Both PWWP and ADD domains, are important for chromatin interactions. The PWWP domain, Pro-Trp-Trp-Pro, is important for DNA binding and it recognizes the H3K9me3 and H3K36me3 marks (Dhayalan et al., 2010; Neri et al., 2017) and it is also required for the localisation to pericentromeric repeats (Ambrosi et al., 2017). The ADD domain, ATRX–DNMT3–DNMT3L, is also important for interactions with chromatin. This domain recognizes unmodified histone H3 and is inhibited by methylation of H3 lysine 4 (Ooi et al., 2009; Otani et al., 2009).

DNMT2 lacks the N-terminal regulatory domain and its catalytic target is not the cytosine nucleotide but it contains all the conserved catalytic motifs of cytosine-5 DNMTs. There is now evidence showing that DNMT2 methylates transfer RNA (tRNA) by employing a DNA methyltransferase-like catalytic mechanism, which suggests that this enzyme changed its substrate specificity from DNA to RNA during its evolution. The tRNA binding site of DNMT2 has been mapped within its catalytic domain (Jurkowski, Shanmugam, Helm, & Jeltsch, 2012).

DNMT3L is a catalytically inactive variant of DNMT3 that lacks the N-terminal part of the regulatory domain which includes the PWWP domain, and the C-terminal part of the catalytic domain, but it maintains the ADD domain (Lyko, 2017).

**Function of DNMT enzymes**

DNMT1 is the most abundant DNMT in somatic cells and its main function is to maintain DNA methylation levels during DNA replication by copying an already established methylation pattern (Yoder, Soman et al. 1997). DNMT1 binds to hemi-methylated DNA at CpG sites after DNA replication. The parental strand remains methylated and DNMT1 catalyses the methylation of the cytosine on the newly synthesized stand. This maintains established CpG methylation patterns through mitosis (Robertson et al., 1999). DNMT1 requires accessory proteins, such as the protein UHRF1, which also plays a role in the maintenance of DNA methylation. UHRF1 preferentially binds to hemi-methylated DNA and is recruited to replication foci to maintain DNA methylation through its interaction with DNMT1 (Ferry et al., 2017).

DNMT3A and DNMT3B establish *de novo* DNA methylation in unmethylated CpGs and do not require hemi-methylated DNA to bind, they show equal affinity for hemi-methylated and non-methylated DNA (Okano, Bell, Haber, & Li, 1999). Both DNMT3A and DNMT3B are required for the genome-wide *de novo* methylation of DNA that occurs during gametogenesis and after embryo implantation in mammalian development. Their activity is facilitated by the ADD domains that enable DNMT3 enzymes to recognise histone H3 N-terminal tail lacking methylation at lysine 4. Regulatory regions in the genome, including CpG islands at gene promoters and enhancers, are protected from methylation by different mechanisms such as bound transcription factors and methylation of histone H3 lysine 4 (Lyko, 2017; Otani et al., 2009). DNMT3A and DNMT3B are highly expressed in undifferentiated embryonic stem cells, but they are downregulated after differentiation with low expression levels in somatic cells (Okano et al., 1999). However, this model where DNMT3 enzymes establish DNA methylation and DNMT1 enzyme maintains the already established patterns of methylation might be oversimplified. There is now evidence showing that there is a crosstalk between *de novo* and maintaining DNA methylation machineries. There are several studies showing that DNMT3A and DNMT3B are also essential for the maintenance of DNA methylation patterns in embryonic stem cells (Chen, Ueda, Dodge, Wang, & Li, 2003) and that the *de novo* DNA methylation is partially catalysed by DNMT1. There are also DNMT-including complexes for *de novo* establishment or maintenance of DNA methylation that can be targeted to particular DNA sequences in a specific or non-specific manner. Specific complexes include the Polycomb proteins and transcriptional factors; while non-specific complexes include heterochromatin readers and replication associated proteins, such as HP1 and UHRF1, respectively (Hervouet et al., 2018). DNMT3B has different isoforms that stimulate gene body methylation independently of their catalytic activity. This suggests a similar functional role to DNMT3L enzyme which recruits DNMT3A to initiate DNA methylation. DNMT3L is only expressed in undifferentiated cells, therefore DNMT3B isoforms might be playing the same role in somatic cells (Duymich, Charlet, Yang, Jones, & Liang, 2016). DNMT3A and DNMT3B have overlapping functions in development but they also exhibit non overlapping functions, since DNMT3B is specifically required for methylation of centromeric minor satellite repeats (Okano et al., 1999) and gene body methylation (Baubec et al., 2015). DNMT3A shows higher DNA methylation activity towards naked DNA and the naked part of nucleosomal DNA while DNMT3B shows activity towards the nucleosome core region, although the activity is low compared to naked DNA. This difference in their activity could contribute to their distinct methylation of genomic DNA regions *in vivo*, which could also be influenced by the difference in timing of expression of DNMT3A and DNMT3B (Takeshima et al., 2006).

DNMT2 is a tRNA methyltransferase that methylates a small set of transfer RNAs at a single specific site (cytosine 38) close to the anticodon loop (Jurkowski et al., 2012). DNMT2 has a post-transcriptional function since methylation of tRNAs protects them against fragmentation having an impact on their stability. The activity of DNMT2 also influences different aspects of protein translation (Lyko, 2017). DNMT3L is an important cofactor for DNMT3 enzymatic activity and increases the affinity of DNMT3A to DNA in undifferentiated cells (Lyko, 2017). Moreover, DNMT3L plays an essential role in the regulation of DNA methylation in gametes, where it guides DNMT3A to establish genomic imprints and facilitates genome stability through the methylation of transposable elements (Ambrosi et al., 2017).

Besides DNMT enzymes, there are other proteins that also affect DNA methylation patterns, such as UHRF1 mentioned earlier and the chromatin remodelling enzymes LSH and ATRX (Dennis, Fan, Geiman, Yan, & Muegge, 2001; Garrick et al., 2000). The role of ATP-dependent chromatin remodelers in DNA methylation remains unclear, but targeting of DNMT3B is impaired and nucleosome occupancy is reduced in the absence of LSH (Ren et al., 2015). The link between DNMTs and human diseases is often discussed in the context of cancer, since aberrant DNA methylation patterns has been shown to affect tumorigenesis. Mutations in DNMTs are rare in human diseases due to the requirement of their activity for appropriate mammalian development. However, ICF syndrome is a developmental genetic disorder characterized by hypomethylation of pericentromeric repeats and is mainly caused by DNMT3B mutations (Jiang et al., 2005), although some patients show mutations in genes that are linked or required for DNA methylation by DNMT3B such as the chromatin remodeler LSH (Thijssen et al., 2015). Another disorder, the Rett syndrome, has been found to result from mutations in the MECP2 gene, which encodes a protein that binds to methylated DNA (Amir et al., 1999).

DNMTs are essential for normal development, but it remains unclear how genome-wide cytosine methylation patterns are established and maintained. Deletion of DNMT3A results in impaired postnatal development while deletion of either DNMT1 or DNMT3B is embryonic lethal in mice indicating a critical role of DNMTs in development (Li, Bestor, & Jaenisch, 1992; Okano et al., 1999).

### 1.3.3 DNA methylation reprogramming during embryonic mammalian development

The patterns of DNA methylation undergo reprogramming during early development (Figure 1.4) which is implicated in early cell lineage specification, genomic imprinting and X chromosome inactivation (Leeb & Wutz, 2012). During early embryonic development, the DNA methylation patterns inherited from the parental gametes are largely erased after fertilization

(P. Zhang et al., 2012). The removal of DNA methylation occurs similarly over the entire genome independently of CpG density, with the exception of imprinted loci and some repetitive elements such as intracisternal A-particle (IAP) elements that maintain high levels of methylation (Reik & Walter, 2001; Smith et al., 2012). The maintenance of DNA methylation at these specific regions requires DNMT1, which is expressed in early embryos from the zygote to the blastocyst stage (Uysal, Akkoyunlu, & Ozturk, 2015). Both parental genomes undergo the same wave of demethylation, but the rate of demethylation is different between the oocyte and sperm. Also, oocytes are already globally hypomethylated compared to sperm (Smith et al., 2012). The paternal genome is subjected to an active and almost complete demethylation after fertilisation and prior to replication (Oswald et al., 2000). However, the maternal genome becomes demethylated passively during cell divisions and it reaches the minimum level of methylation around the blastocyst stage. The majority of the genome at blastocysts stage, before implantation, shows the lowest level of CpG methylation during embryogenesis (Bird 2002, Li 2002). Following this wave of demethylation, DNA methylation is re-established after implantation of the embryo in a wave of *de novo* DNA methylation carried out by DNMT3A and DNMT3B (Okano et al., 1999). However, the level of DNA methylation acquired is different at the embryonic and the extraembryonic lineages. Both lineages gradually gain DNA methylation but the extraembryonic layer contains much less methylation than the embryonic lineages (Chapman, Forrester, Sanford, Hastie, & Rossant, 1984; Monk, Boubelik, & Lehnert, 1987). The *de novo* DNMTs are upregulated at this stage to allow DNA methylation to be re-established genome-wide during the transition to the epiblast stage. The expression patterns and targets of *de novo* DNMTs are different during post-implantation development. DNMT3B might be more important during early development after implantation and methylates a broad spectrum of target sequences, while DNMT3A could methylate specific genes that are critical during late development or after birth (Borgel et al., 2010; Okano et al., 1999). This *de novo* methylation occurs at many types of gene sequences with the exception of CpG islands. There is a correlation between *de novo* methylation and low CpG density regions (Smith et al., 2012). Several promoters become methylated, including pluripotency gene promoters, promoters of gamete specific genes, and numerous promoters associated with tissue specific genes that will be activated later during development (Borgel et al., 2010). Pluripotency genes, such as *Oct4* and *Nanog*, become *de novo* methylated in a specific stage of the development which suggests a role for DNA methylation in the exit from pluripotency (Ambrosi et al., 2017; Borgel et al., 2010).

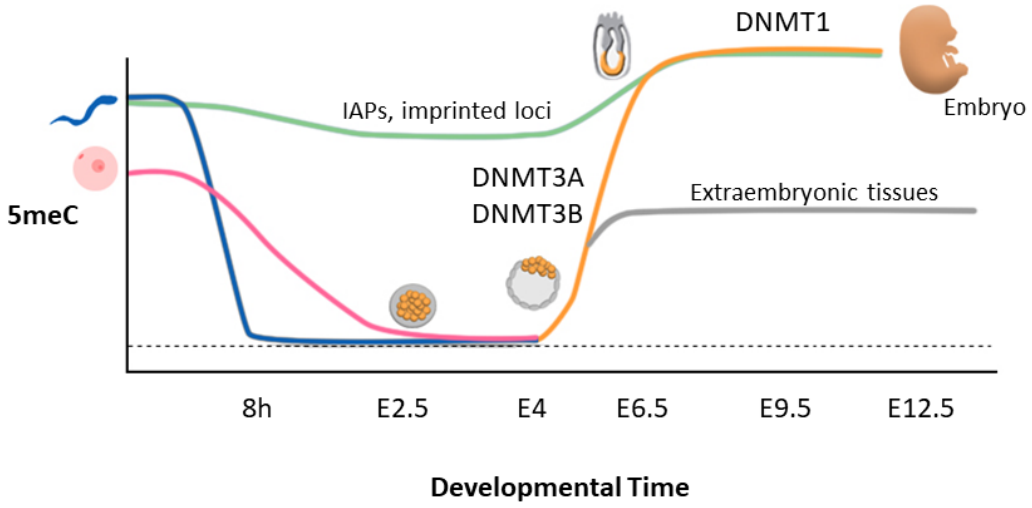


**Figure 1.4. Dynamics of DNA methylation in early mouse development.**

Schematic view of different time points from fertilization in early mammalian development and levels of DNA methylation (5meC) during the different stages of development. Male and female DNA methylation levels are indicated in blue and pink, respectively. During pre-implantation development DNA methylation is lost from both the maternal and the paternal genome. DNA methylation reaches its lowest level at blastocyst stage, before implantation. DNA methylation is re-established after implantation of the embryo in a wave of *de novo* DNA methylation carried out by DNMT3A and DNMT3B enzymes. DNMT1 is responsible for the maintenance of DNA methylation levels. Repetitive IAP elements and imprinted loci maintain high levels of DNA methylation during this process.

The dynamic changes of DNA methylation during development are also accompanied by changes in histone modifications, such as methylation of H3 lysine 4 and lysine 27, that are also essential for development (L. Laurent et al., 2010; Reik, 2007). The importance of post-translational histone modifications will be the focus of the next section.

## 1.4 Chromatin modifications: chemical modification of histone tails

N-terminal tails of histone proteins can be post-translationally modified by different chemical groups such as acetylation, methylation, phosphorylation and ubiquitination. These modifications, which are dynamic and reversible, play an important role as epigenetic factors that in addition to DNA methylation contribute to transcriptional regulation. The histone marks can attract a variety of binding proteins and complexes that can further modify the chromatin architecture and gene expression (Bannister & Kouzarides, 2011; Smith & Meissner, 2013).

The acetylation of lysines in histone tails is regulated by two families of enzymes, histone acetyltransferases (HATs) and histone deacetylases (HDACs). The HAT enzymes catalyse the transfer of an acetyl group, which neutralizes the positive charge of the lysine and weakens the interaction between histones and DNA. Acetylation occurs on numerous lysines in histone tails, including H3K9, H3K14, H3K18, H3K27, H4K5, H4K8 and H4K12. Histone acetylation is enriched at enhancers and gene promoters, where they facilitate the access to transcription factors (Z. Wang et al., 2009). HDAC enzymes remove the acetylation mark, restoring the positive charge of the lysine and therefore stabilizing the local chromatin architecture, which is consistent with HDACs being predominantly involved in transcriptional repression (Bannister & Kouzarides, 2011).

Histone methylation occurs mainly on lysines but also on arginines in histone tails and it does not alter the charge of the histone protein. Lysines can be mono-, di- or tri-methylated and these marks are deposited by different histone lysine methyltransferases (HKMTs). Histone demethylases remove this mark and they are specific to the degree of lysine methylation (Du, Johnson, Jacobsen, & Patel, 2015). The HKMTs usually contain a SET domain that harbours the enzymatic activity. There are different HKMTs, such as the G9a/GLP complex that catalyse mono and di-methylation of H3K9 which is associated with silent genes in euchromatin, and SUV39H1 and SUV39H2 that catalyse the trimethylation of H3K9 a modification mainly associated with centromeric and pericentromeric heterochromatin. SETDB1 is a methyltransferase responsible for the deposition of H3K9me3 mark at endogenous retrovirus and the inactive X chromosome in ES cells and differentiating ES cells, respectively (Du et al., 2015). During embryonic development, the Polycomb group proteins repress gene expression via the formation of complexes containing different subunits referred to as Polycomb repressive complex 1 (PRC1) and 2 (PRC2). PRC2 has histone methyltransferase activity and trimethylates H3K27, a mark of transcriptionally silent chromatin. PRC1 is required for stabilising this silencing through ubiquitylation of histone H2A (A. Bird, 2007; Boyer et al., 2006).

There are other histone modifications such as phosphorylation that reduces the positive charge of histones and ubiquitylation in tails of H2A associated with gene silencing as mentioned earlier (Bannister & Kouzarides, 2011), or ubiquitylation of H3 shown to direct DNMT1 activity at the replication foci (Ferry et al., 2017; Vann & Kutateladze, 2017).

### 1.4.1 Distribution and function of histone modifications

The genome is segmented into different chromatin states with specific chromatin modifications that vary between chromatin types (Filion, van Bemmel et al. 2010). Histone modifications including acetylation or methylation of histone tails reflect either the presence or absence of gene activity in chromatin. While active promoters are characterized by hypomethylated DNA and chromatin methylated at H3K4 residues, silenced chromatin is characterised by low methylation at H3K4 and high methylation levels at H3K9 residue (Bannister & Kouzarides, 2011). However, it is not clear yet whether histone modifications are the cause or consequence of the specific transcriptional state. Induction of H3K4me3 has been shown to induce gene re-expression although this outcome depends on the DNA methylation status of the genomic region (Cano-rodriguez, Gjaltema, Jilderda, Jellema, & Dokter-, 2016).

Transcriptionally active genes are marked by acetylation of histones H3 and H4 in their promoter regions. Histone acetylated lysines are bound by proteins with bromodomain, which are often found in HATs and chromatin remodelling complexes (Bannister & Kouzarides, 2011). In addition to acetylated histones, H3K4me2/3 are also present at active gene promoters, as mentioned earlier. H3K4me3 promotes transcription initiation and H3K9ac could stimulates the release of paused RNA Polymerase II needed for transcription elongation (Gates et al., 2017). Another study showed that Cfp1, the conserved subunit of the Set1 complex that is responsible for H3K4me3, establishes H3K4me3 at promoters upon transcriptional induction and this contributes to H3K9 acetylation, enabling an active chromatin state (Clouaire, Webb, & Bird, 2014). The histone modification H3K4me2 has also been observed in coding regions and correlates with transcriptional activity (Bernstein, Humphrey et al. 2002). Histone modifications not only act as binding sites for protein complexes, they can also inhibit protein interactions such as H3K4me3 that impairs binding of DNMT3 enzymes (Otani et al. 2009).

H3K9 methylation is generally associated with transcriptional silencing and heterochromatin formation, which ensures stable repression and genomic integrity. Gene repression usually correlates with increased H3K9 methylation at promoters and across large genomic regions. The lysine methyltransferases SUV39H1 and SUV39H2 have their main function at constitutive heterochromatin, while SETDB1 and G9a/GLP are responsible for H3K9 methylation at euchromatin and facultative heterochromatin, and are involved in dynamic transcriptional repression (Bannister & Kouzarides, 2011). Repression of transposons and repetitive elements in mouse embryonic stem cells depends on the methylation of histone lysines (Mozzetta, Boyarchuk, Pontis, & Ait-Si-Ali, 2015). The

repression of transcription through H3K9 methylation occurs via the interplay of different enzymes, including DNA methyltransferases, histone deacetylases and H3K4 demethylases (Bannister & Kouzarides, 2011).

Methylation of H3K27 by the PRC2 complex correlates with gene silencing. PRC2 is mainly responsible for the maintenance rather than the establishment of transcriptional silencing and it interacts with other protein complexes, including H3K4 and H3K36 demethylases, to repress transcription through chromatin compaction. This indicates that gene silencing involves removal of activating marks and deposition of repressive H3K27 methylation marks (Mozzetta et al., 2015), which also occurs for other forms of silencing such as methylation of H3K9 that requires removal of active marks. PRC2 is recruited to regions with high CpG content via the zinc finger-CXXC domain of KDM2B that can recognise non-methylated CpGs and favours PRC2 recruitment (Blackledge et al., 2014; Long, Blackledge, & Klose, 2013). However, this recruitment is impaired by transcriptional activity or DNA methylation (Jermann, Hoerner, Burger, & Schübeler, 2014). Developmentally regulated genes in murine embryonic stem cells have been reported to be enriched for bivalent marks that include both active H3K4me3 and repressive H3K27me3 histone modifications. The promoters that contain these bivalent marks in their CpG islands are in an inactive transcriptional state while keeping the genes poised and ready to be activated. Bivalent domains have also been found in intragenic CpG islands that undergo DNA methylation during differentiation. The loss of the bivalent marks results in the activation of their associated genes (S.-M. Lee et al., 2017). The H3K27me3 mark has also been shown to recruit the PRC1 complex that leads to the displacement of PRC2, ubiquitylation of H2A and subsequent gene silencing (Boyer et al., 2006).

### 1.4.2 Histone methylation crosstalk with DNA methylation

Numerous chromatin-associated factors and proteins such as DNMT enzymes interact with modified histones through different domains and produce further changes in chromatin. DNMT enzymes, in addition to their DNA methyltransferase activity, also interact with histone deacetylases and histone methyltransferases. Therefore, DNA methylation is also coupled to transcriptional repression by chromatin modifications (Klose & Bird, 2006). Some interactions between histone methylation and both *de novo* and maintenance DNA methylation have been mentioned earlier, but this section will recapitulate relevant crosstalk mechanisms.

As mentioned earlier, H3K36me3 modification is established in the gene bodies of transcribed genes and associated with transcription elongation. H3K36me3 mark can be recognised by the PWWP domains of the *de novo* DNA methyltransferases and target DNA

methylation to the intragenic region to avoid spurious transcription (Neri et al., 2017; Rondelet, Dal Maso, Willems, & Wouters, 2016). The DNMTs enzymes also contain the ADD domain, which facilitates their interaction with unmethylated H3K4, but this protein interaction is inhibited by the H3K4me3 modification associated with active chromatin at promoters, including unmethylated CpG islands (Otani et al. 2009).

It has been suggested that DNA methylation is linked to the Polycomb protein complexes to stabilize silencing at Polycomb target genes. The H3K27me3 modification mediated by EZH2, component of the PRC2 complex, can lead to recruitment of DNMTs. The link between histone methylation by Polycomb proteins and DNA methylation influences transcriptional repression (Gopalakrishnan, Emburgh, & Robertson, 2008). However, this link is not clear since H3K27me3 and CpG methylation are largely mutually exclusive and DNA methylation has been shown to prevent binding of PRC2 (van Kruijsbergen, Hontelez, & Veenstra, 2015).

Methylation of histone tails and DNA methylation are involved in the formation of heterochromatin, but the relationship between both is still not clear in mammals. The H3K9me3 mark is associated with repressive heterochromatin and there is a specific interaction between this mark and the protein HP1 that recognises the H3K9me3 modification via its N-terminal chromodomain and promotes chromatin condensation (Bannister & Kouzarides, 2011). Also, H3K9me3 has been shown to direct DNA methylation at pericentromeric repeats via the interaction of DNMT3B with HP1 (Lehnertz et al., 2003). Another study shows that transcriptional repression of *Oct-4* gene is followed by an increase in H3K9 methylation that is mediated by the G9a/GLP complex. This leads to local heterochromatinization via the binding of HP1 and is required for subsequent *de novo* methylation at the promoter by DNMT3A/B (Feldman et al., 2006). Another study has shown that G9a recruits DNMT3A/B through its ANK domain, which also interacts with H3 mono- or di-methyl K9 (Epsztejn-Litman et al., 2008). The G9a/GLP complex has also been shown to stabilize imprinted DNA methylation in ES cells by recruitment of *de novo* DNA methyltransferases that antagonize TET-dependent loss of DNA methylation (T. Zhang et al., 2016). These examples indicate the crosstalk between different epigenetic marks and chromatin modifying enzymes.

There are also other proteins involved in the crosstalk between histone modifications and DNA methylation, such as the protein UHRF1 and the ATP-dependent chromatin remodeling protein LSH. It has been shown that UHRF1 plays an important role as a link between histone methylation and maintenance of DNA methylation by DNMT1 at the replication foci. UHRF1 binds to hemi-methylated DNA and methylated DNA ligase 1, which contains a similar motif to H3K9 that is methylated by G9a/GLP (Ferry et al., 2017). Methylated ligase 1 recruits

UHRF1 to replication foci and UHRF1 ubiquitinates H3K23. The ubiquitylation mark recruits DNMT1 via its RFTS domain and stimulates the activity of DNMT1 at the replication foci (Nishiyama et al., 2013). The chromatin remodeler LSH has been associated with pericentromeric heterochromatin and normal CpG methylation at pericentromeric sequences (Dennis, Fan et al. 2001). Loss of LSH results in accumulation of H3K4me2 and H3K4me3 in pericentromeric DNA and other repetitive sequences. This gain of H3K4 methylation follows the loss of CpG methylation caused by LSH deficiency, indicating how LSH is crucial for the formation of normal heterochromatin (Yan, Huang et al. 2003). LSH is also required during embryo development for establishment of normal patterns of DNA methylation genome wide. The loss of methylation observed in the absence of LSH is accompanied by an increase of H3K27me3 at LSH-dependent loci, indicating redundancy of epigenetic silencing mechanisms (Yu et al., 2014). This demonstrates that there is a clear link between DNA methylation, histone modifications, and chromatin remodeling proteins; that contribute collectively to establishment and maintenance of transcriptionally inactive chromatin state.

### 1.5 ATP-dependent chromatin remodelers

Chromatin compaction into higher order structures inhibits the access of many regulatory proteins to DNA, but it also creates an additional layer of organization. Chromatin dynamics change to allow different biological processes to occur and these changes in chromatin organization are predominantly caused by ATP-dependent chromatin remodelers. These proteins are recruited to chromatin through direct binding to epigenetic marks or through proteins that bind epigenetic modifications (Clapier & Cairns, 2009). ATP-dependent chromatin remodelers hydrolyse ATP to obtain the energy required for altering the structure of chromatin (Cairns, 2007). Chromatin remodelers can often be found as part of large remodelling complexes and they lead to different changes in chromatin structure to expose DNA and make it accessible for DNA binding proteins. The different remodeling activities lead to nucleosome sliding through individual nucleosome repositioning, histone octamer eviction that exposes DNA previously wrapped around the histone octamer or nucleosome unwrapping by destabilisation of certain histone-DNA contacts (Figure 1.5). Some chromatin remodelers can alter the histone octamer through either histone dimer eviction or histone dimer exchange, such as H2A-H2B dimer exchange for the histone variant H2A.X-H2B (Clapier & Cairns, 2009).

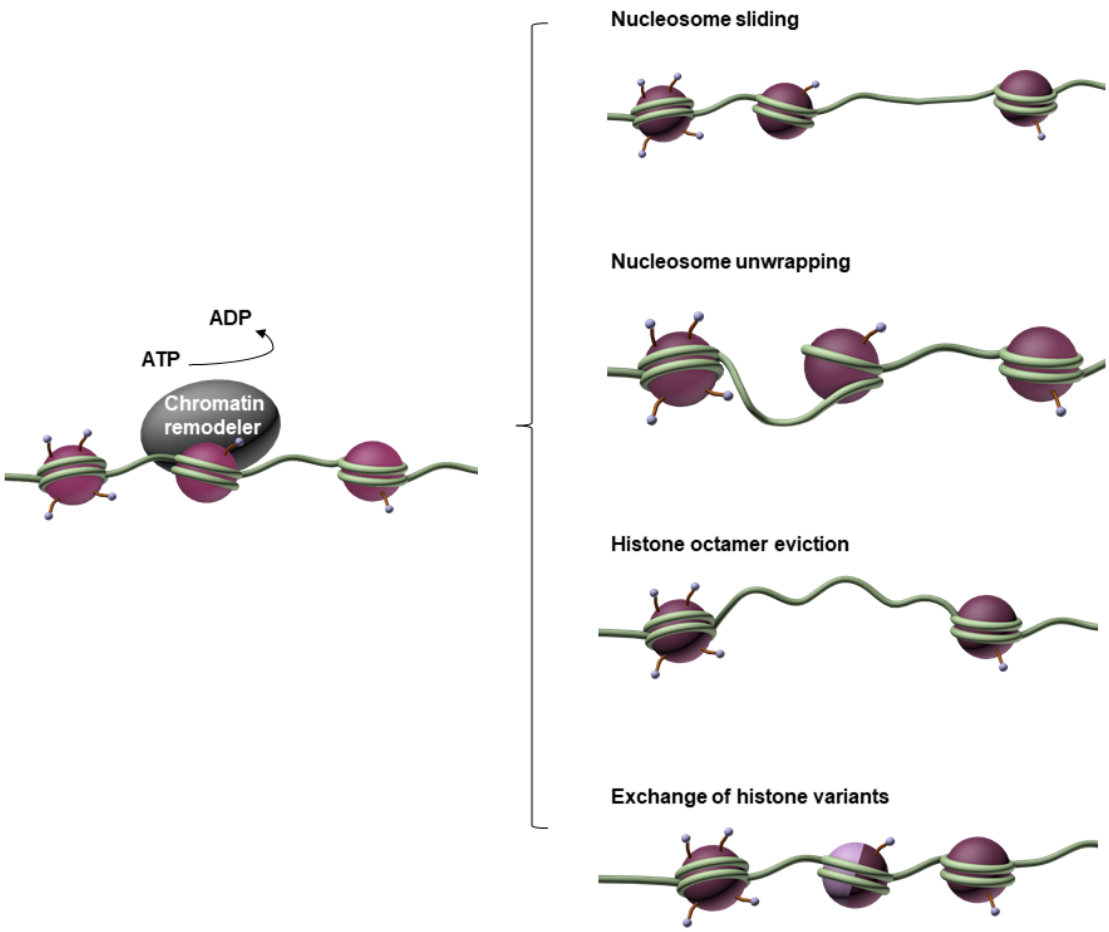


**Figure 1.5. Changes in chromatin structure by ATP-dependent chromatin remodelers or remodeling complexes.**

Three different remodeling activities expose DNA and make it accessible for DNA binding proteins: nucleosome sliding through individual nucleosome repositioning, histone octamer eviction and nucleosome unwrapping by destabilisation of certain histone-DNA contacts. ATP-dependent chromatin remodeling can also alter the histone octamer composition by histone dimer exchange, such as H2A-H2B dimer exchange for the histone variant H2A.X-H2B.

### 1.5.1 Classification of ATP-dependent chromatin remodelers

Chromatin remodeling complexes are divided into four different families based on the similarity of the domains flanking the conserved ATPase region: SWI/SNF, ISWI, CHD and INO80 (Clapier & Cairns, 2009). The four families share a similar ATPase domain allowing them to use ATP hydrolysis to alter the histone-DNA contacts. However, they contain unique additional domains and associated subunits that specialise them for specific functions in

different biological processes. The general domains of chromatin remodelers are the following: a conserved Sucrose Non Fermentable 2 (SNF2) family ATPase domain which couples DNA binding to ATP hydrolysis for DNA translocation, domains that provide affinity to nucleosomes and recognise histone modifications, domains that regulate the ATPase domain and domains that interact with other chromatin regulatory proteins (Clapier & Cairns, 2009). The unique domains are located at flanking regions or between the SNF2 family ATPase regions. These domains provide specialised functions, such as recognition of PTMs, specific transcription factors or chromatin modifiers; that help target remodelers and regulate their function (Clapier & Cairns, 2009; Flaus & Owen-Hughes, 2011). Chromatin remodelers from the SWI/SNF family are characterised by the presence of a bromodomain in the C-terminal region of the ATPase. The bromodomain is a common domain in chromatin remodelers that recognises acetylated lysines in histones and other proteins. The CHD family of remodelers is characterized by the presence of two tandem chromodomains in the N terminal region and PHD finger domains. Chromodomains bind H3K4me2 or H3K4me3 and the PHD finger domain interacts with H3K4me3 and can also recognise core histones rather than the histone tails. The HAND-SANT-SLIDE is a combination of three domains that are found in the C terminal region of the ISWI subfamily and provide nucleosome recognition and stimulation of the ATPase activity. The INO80 subfamily is characterised by an extended linker region that splits the ATPase domain (Clapier & Cairns, 2009).

The empirical classification of ATP-dependent chromatin remodelers into the four families mentioned earlier does not include all the previous subfamilies recognised from a phylogenetic comparison of the SNF2 family ATPase region, such as the subfamilies containing LSH and ATRX. These subfamilies are not included in the previous classification due to their unique flanking regions that cannot be categorised using domain finding tools. A new phylogenetic classification has been made including all known ATP-dependent chromatin remodelers (Figure 1.6). This new classification divides the Snf2 family of chromatin remodelers into 5 groupings and subfamilies that also reflect differences in function (Flaus & Owen-Hughes, 2011). The Snf2 family is named after the first protein of the group to be identified, a member of the Sucrose Non-Fermenter group in *Saccharomyces cerevisiae* (B. C. Laurent, Treitel, & Carlson, 1991).

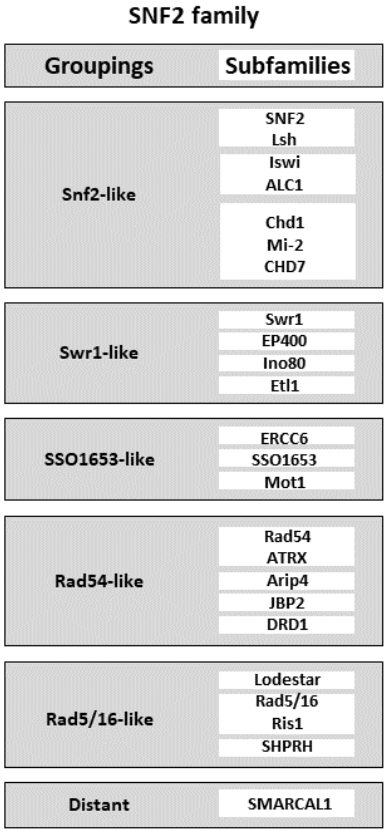


**Figure 1.6. Groupings and subfamilies of ATP-dependent chromatin remodelers defined by phylogenetic comparison of the SNF2 family ATPase region.**

Schematic of the phylogenetic classification of chromatin remodelers into subfamilies. Remodelers shown closer together without gaps are phylogenetically closer than others in the same grouping. Adapted from *Flaus & Owen-Hughes 2011*.

### 1.5.2 Domain architecture and mechanism of action of ATP-dependent chromatin remodelers

Members of the Snf2 family of proteins alter the chromatin state affecting the accessibility of DNA. Several Snf2 factors have been shown to shift nucleosomes along the DNA *in vitro* by disruption of histone-DNA interactions and sliding of nucleosomes along the DNA (Becker & Hörz, 2002; Dennis et al., 2001; Eisen, Sweder, & Hanawalt, 1995; J. Jeddeloh & Bender, 1998), therefore allowing the access of transcription factors and other DNA binding factors to chromatin. ATP hydrolysis is required for the chromatin remodeling function and the ATP binding domain is highly conserved among the Snf2 proteins (Corona et al., 1999).

The ATPase region contains two domains, the SNF2_N and Helicase_C, which contain RecA-like domain lobes 1 and 2, respectively, and are defined by seven conserved helicase motifs (Figure 1.7). The helicase motifs are only required for DNA binding and translocation, they do not unwind double-stranded DNA. The RecA domains are responsible for the ATP hydrolysis, producing the energy necessary for DNA translocation. The ATPase region of Snf2 proteins is characterised by the presence of a structured linker and two antiparallel alpha helical protrusions between the two RecA-like lobes (Flaus, Martin, Barton, & Owen-Hughes, 2006).

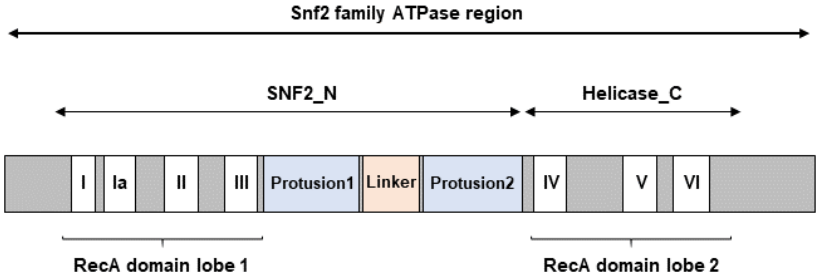


**Figure 1.7. SNF2 family ATPase region architecture and structural features.**

The seven conserved helicase motifs and structural features are shown relative to RecA domain lobes 1 and 2. The figure is not to scale. Adapted from *Flaus & Owen-Hughes 2011*.

Different remodeling mechanisms have been suggested to explain the changes in nucleosome architecture induced by different remodelers, although the different remodeling outcomes might not involve different mechanisms, but modifications in certain kinetic or geometric parameters of the same DNA translocation mechanism (Längst & Becker, 2004). Two main mechanisms have been proposed: twist diffusion and loop propagation. Both models share the idea that a DNA distortion propagates over the surface of the nucleosome.

The twist diffusion model suggests that energy fluctuations could be sufficient to twist the DNA helix at the edge of the nucleosomes, promoting replacement of DNA-histone interactions by 1bp of linker DNA that is pulled into a region between two histone-DNA contacts. This induces increased twist into the double helix, which is subsequently propagated around the nucleosome. This twist diffusion would shift the histone octamer along the DNA by the size of the distortion. The propagation of this twist around the histone octamer surface

would change the rotational and the translational position of the nucleosome (Bowman, 2010; Längst & Becker, 2004).

The loop propagation model suggests that after DNA binds between the RecA domains, structural changes through ATP hydrolysis could result in translocation activity with a segment of DNA being detached at the entry/exit into the nucleosome. The detached DNA fragment may either rebind to form the original nucleosome or interact with a different position on the histone octamer, creating a DNA loop on the nucleosome surface. The loop propagates breaking contacts between histones and DNA at the leading edge of the loop. The propagation of the loop involves the distal linker DNA, resulting in nucleosome repositioning. Depending on the length of the DNA segment detached and the subsequent extent of inclusion of nucleosomal linker DNA into a loop, different size DNA loops may be generated on the nucleosome surface (Cairns, 2007; Clapier & Cairns, 2009; Längst & Becker, 2004).

In addition to the previous models, an alternative histone core swivel model has been proposed. This model involves rotation of the histone core with respect to the wrapping DNA and subsequent realignment of the DNA around it. The remodeler binds two helical turns away from the DNA entry site of the nucleosome. ATP hydrolysis by the remodeler is accompanied by a DNA conformational change, which facilitates the histone core to move one helical turn towards the remodeler and then DNA rewraps around (Bowman, 2010).

### 1.5.3 Functional diversity of ATP-dependent chromatin remodelers

The Snf2 family of chromatin remodelers, contains proteins which play important biological functions including DNA repair (ERCC6/CSB, INO80); genomic recombination (RAD54); transcriptional control (SNF2); chromosome segregation and checkpoint control in DNA replication (INO80); and chromatin assembly to maintain higher order structures (ISWI) (Eisen et al., 1995; Ho & Crabtree, 2010). The members of this family can be found as a subunit of chromatin remodeling complexes, which explains the functional diversity of these enzymes (Clapier & Cairns, 2009). The Isw2 complex has been shown to increase nucleosome occupancy of the intergenic region to prevent antisense transcription in *S. cerevisiae* (Whitehouse, Rando, Delrow, & Tsukiyama, 2007). Most ISWI complexes promote the equal spacing of DNA between nucleosomes to form periodic nucleosome arrays (Cairns, 2007). CHD3 and CHD4 are subunits of the NURD (nucleosome-remodelling and histone deacetylase) complex, which contains histone deacetylases and functions as a transcriptional repressor. Mutations in CHD7, a chromatin remodeler involved in transcriptional activation of tissue-specific genes during differentiation, result in CHARGE syndrome which is a developmental syndrome in humans (Ho & Crabtree, 2010). The function of the Snf2 family

proteins can be quite complex. Human patients null for the Snf2 member ATRX (α-thalassemia/mental retardation syndrome X-linked homolog) have methylation alterations in DNA sequences at highly repeated sequences including a Y-specific satellite and subtelomeric repeats. These patients also exhibit a variety of developmental defects, including mental retardation and facial dysmorphology (Garrick et al., 2000). The homologs LSH (*Mus musculus*) and DDM1 (*Arabidopsis thaliana*) form their own subclass within the Snf2 family and are known to regulate DNA methylation during development (Dennis et al., 2001). These findings demonstrate that chromatin remodelers are essential during development and some Snf2 members also play a role in DNA methylation (Ho & Crabtree, 2010; Raabe, Abdurrahman, Behbehani, & Arceci, 2001).

## 1.6 LSH, an ATP dependent chromatin remodeler

LSH (lymphoid-specific helicase), also known as HELLS (helicase lymphoid-specific) or PASG (proliferation associated SNF2-like gene), belongs to the Snf2 family of proteins (Jarvis, Geiman et al. 1996), whose members participate in ATP-dependent chromatin remodeling (Längst and Becker 2004). Previous studies identified LSH as a transcript highly expressed in lymphocytes, leading to the proposed name of Lymphocyte Specific Helicase (LSH) (Jarvis et al., 1996). LSH was found to be expressed only in thymus, early lymphocytes and activated lymphocytes, but it has been shown early on that LSH is highly expressed throughout the embryo in early development (Raabe et al., 2001). LSH is expressed ubiquitously, but its expression is most prominent in proliferating tissues (D. W. Lee et al., 2000; Raabe et al., 2001). A recent publication has shown that LSH forms part of a bipartite nucleosome remodeling complex, which includes a relatively understudied Myc-regulated zinc finger protein CDCA7. LSH alone fails to remodel nucleosomes, but CDCA7 stimulates the nucleosome remodeling activity of LSH. CDCA7 also facilitates binding of LSH to chromatin, although CDCA7 mutants defective in recruiting LSH can still support its remodeling activity. This suggests that CDCA7 can activate LSH by an allosteric mechanism in addition to its role in chromatin recruitment (Jenness et al., 2018).

### 1.6.1 Domain architecture of LSH

Murine *Lsh* gene is localized on chromosome 19 (the human homologue on chromosome 10) and is composed of 24 exons. The seven helicase motifs that define the ATPase domain are distributed on seven exons (T M Geiman, Durum, & Muegge, 1998) and there is 99% homology of the helicase motifs between mouse and human LSH. The difference between

human and mouse LSH is the presence of 16 additional amino acids in the N-terminal region of the human protein (Briones & Muegge, 2012). Members of the Snf2 family, which are involved in transcriptional control and DNA repair, usually contain a chromodomain, bromodomain or a ring finger motif, but none of these are present in LSH (Eisen et al., 1995). The functional domains of murine LSH protein are shown in a schematic diagram in Figure 1.8.

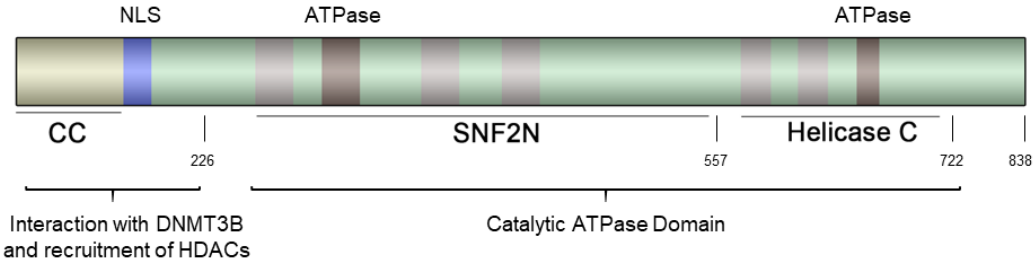


**Figure 1.8. Schematic view of murine LSH protein domains.**

CC (Coiled Coil) motif and NLS (Nuclear localization signal) map to the N-terminal region (amino acids 1-226) of the protein. The N-terminal region is required for interaction with DNMT3B and recruitment of HDAC1/2. The catalytic ATPase region includes the SNF2_N and Helicase_C domains with seven helicase motifs (from amino acid 226 to amino acid 722), and belongs to the conserved motif in the Snf2 family.

The N-terminal region of LSH (amino acids 1-226) contains a coiled-coil (CC) region and a nuclear localisation signal (NLS) (D. W. Lee et al., 2000). The N-terminal region of LSH seems to be necessary for protein binding since co-immunoprecipitation experiments have shown that the CC domain of LSH binds to E2F3 (von Eyss et al., 2012) and the N-terminal region of LSH interacts with DNMT3B, that indirectly recruits HDAC1/HDAC2 (Myant and Stancheva 2008). The Snf2 ATPase region (from amino acid 226 to amino acid 722) contains the SNF2_N and Helicase_C domains with the seven helicase motifs. This ATPase region is highly conserved with other Snf2 family members (Eisen et al., 1995).

### 1.6.2 Role of LSH in DNA methylation

Screening in *Arabidopsis thaliana* for DNA methylation mutants, that show genome hypomethylation, identified the protein Decrease in DNA methylation 1 (DDM1) (Vongs, Kakutani, Martienssen, & Richards, 1993). There is a 70% reduction in DNA methylation levels in DDM1-deficient plants, with repetitive elements and single copy genes showing hypomethylation. Moreover, DDM1-dependent methylation is required for gene silencing in *Arabidopsis* (J. Jeddeloh & Bender, 1998; Vongs et al., 1993). DDM1 was found to be a SNF2-like protein by sequence similarity (J. A. Jeddeloh, Stokes, & Richards, 1999) and the ATP-dependent nucleosome repositioning activity of DDM1 was also confirmed (Brzeski & Jerzmanowski, 2003). Therefore, DDM1 was the first Snf2 family chromatin remodeling protein found to be involved in DNA methylation (J. A. Jeddeloh et al., 1999). LSH, the mammalian homolog of DDM1, was identified in early developing lymphocytes in a study attempting to find helicases involved in VDJ recombination (Jarvis et al., 1996). Later on the role of LSH in DNA methylation was studied due to its sequence conservation with DDM1 (J. A. Jeddeloh et al., 1999) and it was found that, similarly to DDM1, LSH is required for genome-wide DNA methylation (Dennis et al., 2001).

Two different design strategies have been used to generate mice deficient in LSH in order to study the biological functions of LSH. Kathrin Muegge laboratory generated knockout mice through a targeted deletion of some exons in the *Lsh* gene using homologous recombination. The exons 6 and 7 with helicase domains I, Ia (ATPase domain) and part of domain II were removed; and the protein translation prematurely interrupted at the deletion. No wild-type or truncated forms of LSH protein were found in knockout mice (T M Geiman et al., 2001). The second strategy consisted on the disruption of *Lsh* gene through homologous recombination by deletion of the conserved helicase domains II, III, IV (exons 10-12). The deletion of these exons was shown to truncate the catalytic SNF2 domain generating a hypomorph *Lsh* allele. A truncated form of mRNA with predicted length was observed in the mice carrying this deletion. However, a short form of LSH protein with the deletion of 152 amino acids was detected in immunoprecipitation experiments, but not by Western blots, indicating that the expression of the truncated protein was very low (Sun et al., 2004).

These two knockouts have enabled to study the role of LSH in different biological processes. LSH is crucial for cell proliferation and normal development, and targeted deletion of LSH in mice leads to early postnatal lethality (T M Geiman et al., 2001; Sun et al., 2004). LSH has also been implicated in the control of CpG methylation (Dennis et al., 2001). It has been reported that the acquisition of DNA methylation at some loci in the genome depends on the presence of LSH, while the maintenance of previously methylated DNA does not require LSH (Zhu et al., 2006). However, our observation using mouse embryos (unpublished data)

suggests that LSH might support maintenance as well as *de novo* DNA methylation. LSH deficient cells and embryos show reduced DNA methylation by ~50% in cells and by ~40% in embryos (Myant & Stancheva, 2008; Tao et al., 2011), local increased levels of H3K4me3 (Tao et al., 2011; Termanis et al., 2016), and local increased acetylation levels for histone H3 and H4 (Dennis et al., 2001; Termanis et al., 2016). LSH is highly expressed in pluripotent embryonic stem cells and shows reduced expression in somatic cells and adult tissues. The effect of LSH on stem cell gene expression was shown to involve differentiation-dependent DNA methylation (Dennis et al., 2001). Furthermore, LSH is crucial for the control of heterochromatin at pericentromeric regions consisting of satellite repeats. The absence of LSH causes an increased association of acetylated histones with repeat sequences and de-repression of retrotransposons, indicating an important function of LSH protecting against deregulation of retroviral elements in the genome (De La Fuente et al., 2006; T. Fan, Hagan, Kozlov, Stewart, & Muegge, 2005; Huang et al., 2004). LSH is required for DNA methylation and appropriate gene expression not only at repetitive elements, such as centromeric and pericentric satellite repeats, IAP retrotransposons, and LINE elements, but also at protein coding genes (Myant et al., 2011). More recent studies have also shown that specific protein coding genes are affected by lack of LSH as well as repetitive sequences and large chromosomal domains associated with the nuclear lamina (Myant et al., 2011; Yu et al., 2014).

LSH associates with DNMT3A or DNMT3B, but not with DNMT1, in embryonic cells (Zhu et al., 2006). Other study has also shown that LSH interacts with DNMT3B and indirectly recruits DNMT1 and HDAC1/HDAC2. The N-terminal region of LSH is required for these interactions and sufficient for HDAC-mediated silencing (Myant & Stancheva, 2008). LSH has been shown to be necessary, but not sufficient, to control DNMT3B activity at specific genomic sites in ES cells (Xi et al., 2009). In relation to this, LSH plays a role in transcriptional silencing of the developmentally regulated *Hox* genes (Xi et al., 2007). However, not all *de novo* DNA methylation events are LSH dependent. Many promoters that are methylated in wild-type mouse embryonic fibroblasts (MEFs) have normal levels of DNA methylation in *Lsh* knockout MEFs, while 20% of the promoters show loss of DNA methylation in the absence of LSH (Myant et al., 2011). Extensive DNA methylation and expression defects in *Lsh* knockout MEFs has been characterised genome-wide, showing that LSH is directly involved in regulating DNA methylation throughout the genome at single gene promoters as well as clusters of promoters and repetitive sequences (Myant et al., 2011). LSH can partly overcome these defects cell autonomously, when introduced into the *Lsh* knockout MEFs. However, the catalytic activity of LSH is necessary for re-establishment of DNA methylation and transcriptionally silent state (Termanis et al., 2016). Supporting these experiments, a recent publication has shown that the ATPase function of LSH is critical for *de novo* methylation at repeat sequences. It was also

suggested in this study that ATP-dependent nucleosome remodeling is the primary molecular function of LSH, which may promote *de novo* methylation in differentiating ES cells (Ren et al., 2015).

In summary, LSH is essential for developmentally programmed *de novo* DNA methylation at the promoters of protein coding genes (Myant et al., 2011) and also controls genome-wide cytosine methylation at repetitive and unique sequences (Tao et al., 2011). How DNA methylation at LSH-dependent loci is established is currently unclear. One hypothesis is that part of the LSH effect may be due to the stabilization of DNMT3B association with specific genomic loci since LSH depletion reduces DNMT3B binding to specific sites at *Hox* genes in ES cells (Xi et al., 2009). However, it still remains unclear whether LSH is acting directly or indirectly to control DNA methylation. Furthermore, it is not well understood whether or not chromatin remodeling is required for LSH function or if LSH plays a role as a scaffolding protein to promote the association of the DNA methylation machinery with chromatin-organised DNA. Finally, the dynamics and the order of events that lead to LSH-dependent DNA methylation are also unclear and have never been investigated in detail.

### 1.6.3 Involvement of LSH in other biological processes and disease

#### LSH in DNA repair

Previous experiments in yeast (Alvaro, Lisby, & Rothstein, 2007), which carry an LSH homologue but entirely lack DNA methylation, and *Arabidopsis thaliana* (Shaked, Avivi-Ragolsky, & Levy, 2006) suggested that LSH may have an additional function in DNA repair that is independent of its role in DNA methylation and regulation of gene expression. This inspired the study in mammalian cells and LSH was shown to be necessary for the repair of DNA double-strand breaks (DSBs), which is independent of its function in *de novo* DNA methylation. LSH-deficient mouse and human fibroblasts show reduced viability after exposure to ionizing radiation and reduced efficiency in repairing DSBs compared to wild-type cells (Burrage et al., 2012). LSH-deficient cells show normal activation of the DNA damage-responsive kinase ATM, which initiates the DNA damage response signalling, but weaker and more transient phosphorylation of the ATM substrate H2A.X. Phosphorylated H2A.X facilitates the recruitment of DNA damage repair mediator proteins, which localise at ionizing radiation foci to initiate DNA repair, and coordinates cell cycle arrest during double-strand break repair. However, H2A.X is not efficiently phosphorylated in response to DNA damage in the absence of LSH and therefore there is reduced recruitment of DNA damage response proteins, which

leads to inefficient DNA repair and apoptosis of LSH-deficient cells. DSB repair defects in LSH-deficient cells are independent of changes in DNA methylation and the ability of LSH to hydrolyse ATP is necessary for efficient phosphorylation of H2A.X at DSBs and efficient repair of DNA damage (Burrage et al., 2012).

In relation to this, MUS-30, the LSH homolog in the filamentous fungus *Neurospora crassa*, has been reported to be required for genome stability but not DNA methylation. MUS-30 deficient cells have normal DNA methylation, but are hypersensitive to DNA damaging agents. MUS-30 was shown to prevent DNA damage that arises from toxic base excision repair intermediates, which indicates that the LSH homolog has a function in DNA repair that is independent of DNA methylation. The function in DNA repair is most likely the ancestral function of the protein as it is conserved from *S. cerevisiae* to plants (Shaked et al., 2006) and humans (Basenko, Kamei, Ji, Schmitz, & Lewis, 2016; Burrage et al., 2012). However, it is not clear whether deficient DNA repair can explain the meiotic defects observed in $Lsh^{-/-}$ germ cells (De La Fuente et al., 2006) and premature ageing phenotype in mice expressing hypomorph alleles of LSH (Sun et al., 2004).

**LSH in cancer**

LSH has been found to be involved in different cancers, such as leukaemia (D. W. Lee et al., 2000), head and neck squamous cell carcinoma (Waseem, Ali, Odell, Fortune, & Teh, 2010), prostate cancer (von Eyss et al., 2012) and lung adenocarcinoma (R. Wang et al., 2015). LSH overexpression significantly correlates with poor survival in patients with lung adenocarcinoma cancer (R. Wang et al., 2015) and LSH expression positively correlates with progression of head and neck squamous cell carcinoma (Waseem et al., 2010), most likely indicating the high proliferation capacity of these cancers.

The predominant isoform of p63 transcription factor is a protein that enables cells to bypasses oncogene-induced senescence and promotes stem-like proliferation and carcinoma development. LSH is a direct target of this p63 isoform in primary keratinocytes and LSH knockdown in keratinocytes causes increased senescence (Keyes et al., 2011). This study shows that p63-mediated LSH expression is essential to bypass senescence and promote tumorigenesis in keratinocytes (Keyes et al., 2011).

The function of the pRB (retinoblastoma) tumour suppressor protein was originally thought to be exclusively due to its capacity to arrest cells in G1 by inhibiting the activity of E2F transcription factors. In human patients, aberrant E2 factor transcription factor 3 (E2F3) expression is linked to different types of cancer (Burkhart & Sage, 2008). LSH has been

identified as an interacting partner of E2F3. LSH was shown to bind E2F3 and to promote induction of E2F target genes and re-entry into the cell cycle. Both, LSH and E2F3 are overexpressed in several human tumours and they act as markers of aggressive stage in prostate carcinomas (von Eyss et al., 2012). LSH binds 93% of the transcriptional start sites bound by E2F3 and 86% of these promoters contain H3K4me3, a mark associated with active gene transcription. This study shows that LSH positively co-regulates E2F3-dependent transcription and is critical for cell proliferation in cancer (von Eyss et al., 2012).

**LSH in ICF syndrome**

Mutations in LSH have been found in patients with Immunodeficiency–Centromeric instability–Facial anomalies (ICF) syndrome (Thijssen et al., 2015). This syndrome is characterized by centromeric instability and CpG hypomethylation of centromeric satellite repeats. ICF syndrome is normally caused by mutations in the catalytic domain of DNMT3B that explains 50% of cases and mutations in the zinc-finger and BTB domain containing 24 (ZBTB24) gene that explains 30% of cases (De Greef et al., 2011; Jiang et al., 2005). A recent study has found mutations in LSH and cell division cycle associated 7 (CDCA7) genes in ten unexplained cases of ICF. Five point mutant variants of LSH were identified in patients with ICF syndrome (Thijssen et al., 2015). The role of LSH in the establishment of DNA methylation shown in previous studies (Zhu et al., 2006) can explain its involvement in ICF syndrome. However, it was not clear how ZBTB24 and CDCA7 are involved in ICF syndrome, but recent studies have found that ZBTB24 can function as a transcription factor directly controlling CDCA7 expression (Wu et al., 2016) and CDCA7 was recently shown to form a bipartite nucleosome remodeling complex with LSH (Jenness et al., 2018). CDCA7 helps recruit LSH to chromatin and stimulates its remodeling activity. Notably, CDCA7 mutants in ICF syndrome are defective in chromatin recruitment of the LSH-CDCA7 complex. Therefore, ZBTB24 and CDCA7 mutations in patients with ICF syndrome may cause defects in LSH recruitment to chromatin and stimulation of LSH remodeling activity, affecting DNA methylation (Jenness et al., 2018).

## 1.7 Aims

My project focuses on the study of the role of LSH in the establishment of epigenetic gene silencing. I aimed to investigate how LSH-mediated changes in chromatin structure and modifications facilitate *de novo* DNA methylation and gene silencing. LSH-null mice carrying conditionally reversible *Lsh* alleles were used to obtain mouse embryonic fibroblast (MEF) cell lines for the study. I characterised MEFs containing conditionally reversible *Lsh* knockout alleles and optimised the conditional expression of LSH in this system. The specific aims of my project were to address the following questions:

- What is the efficiency of LSH expression in the conditional knockout cellular system?

- How much time is required for acquisition of gene silencing and DNA methylation at the promoters of misregulated loci after restoration of LSH expression?

- What is the order of events leading to gene silencing and *de novo* DNA methylation?

- What is the mechanism of LSH-dependent gene silencing?

# Chapter 2 - Materials and Methods

## 2.1 Materials

### 2.1.1 Antibodies

**Western Blotting antibodies**

All antibodies used at the indicated working concentration in 4% milk + 0.1% Tween 20.

- LSH: Mouse monoclonal Santa Cruz sc-46665, 1:1000 dilution.
- HDAC1: Rabbit polyclonal Santa Cruz sc-7872, 1:500 dilution.
- H3Ac: Anti-acetyl histone H3 rabbit polyclonal IgG Upstate Millipore 06-599, 1:3000 dilution. Acetyl-histone H3 (Lys9, Lys 14) rabbit polyclonal Invitrogen PA5-16194, 1:1000 dilution.
- H3K9me2: Mouse monoclonal Abcam ab1220, 1:2000 dilution.
- Histone H4: Pan rabbit monoclonal Millipore 04-858, 1:1500 dilution.
- Secondary antibodies: Donkey anti-mouse IR 800 or donkey anti-rabbit IR 680LT and IR 800, LI-COR Biosciences, 1:10000 dilution.

**ChIP antibodies**

4 micrograms of antibody used per IP reaction unless otherwise indicated.

- Histone H3: Rabbit polyclonal. Abcam ab1791.
- H3Ac: Anti-acetyl histone H3 rabbit polyclonal IgG Upstate Millipore 06-599.

  Acetyl-histone H3 (Lys9, Lys 14) rabbit polyclonal Invitrogen PA5-16194, 1:50 dilution.
- H3K4me3: Rabbit polyclonal Active Motif 39915.
- H3K9me2: Mouse monoclonal Abcam ab1220.
- H3K9me3: Rabbit monoclonal to Histone H3 (try methyl K9) Abcam ab 176916.
- H3K27me3: Tri-Methyl-Histone H3 (K27) (C36B11) Rabbit. Cell Signalling Technology.
- LSH: Anti-lymphoid-specific helicase (HELLS) rabbit polyclonal. Millipore ABD41.
- Control mouse IgG and control rabbit IgG (Sigma).

### 2.1.2 Cell lines

- Wild-type immortalised MEFs.
- *Lsh*$^{-/-}$ immortalised MEFs: Isolated from *Lsh*$^{-/-}$ mouse embryos derived from crosses of heterozygous animals (Kathrin Muegge).
- *Lsh*$^{-/-}$ MSCV, *Lsh*$^{-/-}$ LSH and *Lsh*$^{-/-}$ K237Q: *Lsh*$^{-/-}$ MEFs were infected with packaged lentiviral particles carrying pMSCV, pMSCV-LSH and pMSCV-K237Q plasmids (Ausma Termanis). All are immortalised cell lines.
- *Lsh*$^{+/+}$ *ER-Cre*$^{+}$, *Lsh*$^{off/off}$ *ER-Cre*$^{+}$ and *Lsh*$^{on/on}$ *ER-Cre*$^{+}$ MEFs: primary cell lines isolated from E13.5 stage mouse embryos derived from crosses of heterozygous animals for the *Lsh*$^{off}$ and *Lsh*$^{on}$ alleles. Some of these cell lines were immortalised.

### 2.1.3 Primers

Primers were designed using DNASTAR Lasergene software and were purchased from MWG Eurofins. Lyophilised primers were diluted in distilled $H_2O$ to 100 µM stock concentration and stored at -20°C. The working concentration of the primers is 5 µM. See following tables for sequences of primers.

**PCR Primers:**

| | |
|---|---|
| eGFP-Neo (*Lsh*$^{off}$) | CGCCGACCACTACCAGCAGAACA |
| | GCCGGAGAACCTGCGTGCAATC |
| Lsh WT | CCTCCCCAAATAAGCAAATAAAAACT |
| | CGAAGGTTGCCAGGTTTTGAGATC |
| Lsh on/off | CGCCTTCTATCGCCTTCTTGAC |
| | CAAGTTCAGCGTGTCCGGCG |
| | GCCATAAGCCAACCAACAGAAAGAATAAG |
| SRY | TTGTCTAGAGAGCATGGAGGGCCATGTCAA |
| | CCACTCCTCTGTGACACTTTAGCCCTCCGA |

**qPCR Primers:**

| | |
|---|---|
| GAPDH | TGGTATCGTGGAAGGACTCATGAC |
| | ATGCCAGTGAGCTTCCCGTTCAGC |
| Lsh off | GCTTCTGAGACGGAAAGAACCAG |
| | GCGGAAGAGTCTAGAATAACTTCGTATAG |

**qRT-PCR Primers:**

| | |
|---|---|
| Dppa2 | GCA TTC ATT CAG CGG CTG CCT TT |
| | TGC GTA GCG TAG TCT GTG TTT GG |
| Dppa4 | CAA GGG CTT TCC CAG AAC AAA TGC |
| | GCA GGT ATC TGC TCC TCT GGC AC |
| GAPDH | TGGTATCGTGGAAGGACTCATGAC |
| | ATGCCAGTGAGCTTCCCGTTCAGC |
| Gm9 | GCTGAATCCGTTGGGCACAGTGAA |
| | GGAGGTAAAGGTCGTTGGGAGCCA |
| Hoxa4 | AAGATCCACGTGAGCGCCGTCACT |
| | GCTCCGACAAGCAGAGCGTGTG |
| Hoxa5 | CACATCAGCAGCAGAGAGGGGGTT |
| | TCGGGCCACCTATATTGTCGTGACT |
| Hoxc6 | GAATGAATTCGCACAGTGGGGTCG |
| | TTTGATCTGTCGCTCGGTCAGGCA |
| IAPs | ACTAACTCCTGCTGACTGG |
| | TGTGGCTTGCTCATAGATTAG |
| MuLV | GGCGCCCCGTACAAGATTTCATA |
| | GATAACGGGCCTGCCTTCACCTC |
| Ndn | CTTCCCGATACAGGGAGGCTCTGG |
| | CAACCCCACCCTTACACAGCCCTA |
| Peg3 | CAGAGGACCCTGACAAGGAG |
| | AGCACAGCACTCTACGCACA |
| Peg12 | TGTTGTAGTAGGAGGAAGTTGT |
| | CTTAACCTCCACCCAACCTAAA |

| | |
|---|---|
| Rhox2 | CACTAAGGAGGCAAATCGCCTGGCA |
| | ACAGCCTTCACAGAGAGATGCTCCG |
| Rhox6 | CAGCTTGCGAGCAAGGAGGGATCT |
| | ATCTTCAGGGAGAGTCGCTCTGGG |
| SINE B1 | GTGGCGCACGCCTTTAATC |
| | GACAGGGTTTCTCTGTGTAG |
| SINE B2 | GAGATGGCTCAGTGGTTAAG |
| | CTGTCTTCAGACACTCCAG |
| Tnfsf9 | ATTCACAAACACAGGCCACAA |
| | AACGGTCCACTAACTTGTTCTC |

**ChIP qPCR Primers:**

| | |
|---|---|
| ActinB | ATGAAGAGTTTTGGCGATGG |
| | GATGCTGACCCTCATCCACT |
| Dppa2 | AAAAGAAGTCGGCATTCATTCAGC |
| | AGCCCTAGAGTGAAAAAGAACTCC |
| Gm9 | TTTCTCTATAGCAGCCCAAGTACC |
| | AAGCGCCTAAAGTTTCAAAGACAC |
| Rhox2 | GAATAAGGACTTCCACGGCTTTAC |
| | TCCAGTTCAATTCTATATGGGATGC |
| Rhox6 | CCATGTTGCTCAGGTCTTTATCTC |
| | GAGCGAGCCAGTTCAGTACAAG |
| Rhox9 | GAGCAAGCCCGCTCATAACAAG |
| | GCCATGATGCTCAGATCTTTATCTC |
| Tnfsf9 | CTCTATGGCCTAGTCGCTTTGG |
| | GTGACCTGGTCTGCATTATTCTC |
| IAPs | CAAATAATCTGCGCATATGCCGA |
| | CAGGAGCAAGAGTGTAAGAAGC |

**Primers for Bisulfite-converted DNA amplification:**

| | |
|---|---|
| Dppa2 | TGTTATGTTTTGTTTTGGATTTAAAA |
| | AAAAAAACCAATTCTCTTAAAAAATCA |
| Dppa2 nested | TGTTATGTTTTGTTTTGGATTTAAAA |
| | TCCTATAACCCAATTAAACTCTTCC |
| Dppa4 | AGTTGTTAGGAGTAGGGGGTAGTAGT |
| | AAAACCCTCATTTAAAAAACCAAAT |
| Dppa4 nested | AGTTGTTAGGAGTAGGGGGTAGTAGT |
| | TCCAACAATCTCCATCTTAAAAATAA |
| Gm9 | GGGGGTATATTTTTTGAAATTTTTTAG |
| | CCACTATCCTACCTATCACAACCAT |
| Gm9 nested | GGGGGTATATTTTTTGAAATTTTTTAG |
| | CATCCAATCACAAACCTTAAACATA |
| Rhox6/9 | GTTTGAGGAGGTTTTTGTGTTATT |
| | AAATCCAATCAAAACCAATTTACT |
| Rhox6/9 nested | TTATTTTTTGTGAAGGAATTTGAA |
| | AAATCCAATCAAAACCAATTTACT |
| IAP | TTGATAGTTGTGTTTTAAGTGGTAAATAAA |
| | AAAACACCACAAACCAAAATCTTCTAC |
| IAP nested | TTGTGTTTTAAGTGGTAAATAAATAATTTG |
| | CAAAAAAAACACACAAACCAAAAT |

**Sequencing Primers:**

| | |
|---|---|
| pJET Seq-pJET-F | GCCTGAACACCATATCCATCC |

### 2.1.4 Buffers and solutions

**General buffers**

Buffers and reagents were kept at room temperature unless otherwise stated.

- Tris-EDTA (TE): 10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.
- TAE (Tris-acetate EDTA): 40 mM Tris, 20 mM glacial acetic acid, 1 mM EDTA pH 8.
- PCR buffer IV: 750 mM Tris-HCl pH 8.8, 250 mM $(NH_4)_2SO_4$, 0.1% Tween 20, 25 mM $MgCl_2$.
- Special PCR buffer for Bisulfite Sequencing: 166 mM (NH4)2SO4, 670 mM Tris pH 8 and 100 mM Beta-mercaptoethanol.
- 2.5 X Sequencing buffer: 20 mM Tris-HCl pH 8, 5 mM MgCl2.
- Bisulfite conversion solution: 3.8 g sodium bisulfite was dissolved in 5 ml water and 1.5 ml freshly made 3 M NaOH in the dark. 110 mg hydroquinone was dossilved in 1 ml water at 55°C for 10 minutes and subsequently added to the sodium bisulfite solution.
- SOC media: 20 g/L Difco Bacto tryptone, 5 g/L Difco Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4, 20 mM glucose.
- LB media: 10 g/L Difco Bacto tryptone, 5 g/L Difco Bacto yeast extract, 5 g/L NaCl, pH 7.2.
- LB Agar: LB media + 2% Difco Bacto agar.

**ChIP buffers**

- Crosslinking solution (10x): 500 mM Hepes pH 7.6, 1.5 M NaCl, 10 mM EDTA, 5 mM EGTA.
- Buffer L1 for ChIP: 50 mM Hepes pH 7.9, 140 mM NaCl, 1 mM EDTA, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100.
- Buffer L2 for ChIP: 10 mM Tris pH 8, 200 mM NaCl, 1 mM EDTA, 0. 5mM EGTA.
- Buffer L3 for ChIP: 10 mM Tris pH 8, 1 mM EDTA, 0.5 mM EGTA.
- ChIP dilution buffer: 20 mM Tris pH 8, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100.
- Wash buffer 1 for ChIP: 20 mM Tris pH 8, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS.

- Wash Buffer 2 for ChIP: 20 mM Tris pH 8, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS.

**Protein extraction and western blotting buffers**

- NE1: 20 mM Hepes pH 7.0, 10 mM KCl, 1 mM MgCl2, 0.1% (v/v) Triton X-100, 20% (v/v) glycerol. 0.5 mM DTT and complete protease inhibitors (Sigma) were added at time of use.
- NE2: 20 mM Hepes pH 7.0, 500 mM NaCl, 10 mM KCl, 1 mM MgCl2, 0.1% (v/v) Triton X-100, 20% (v/v) glycerol. 0.5 mM DTT and complete protease inhibitors (Sigma) were added at time of use.
- Ponceau S staining solution: 1% (v/v) glacial acetic acid, 0.5% (w/v) Ponceau S.
- SDS PAGE loading buffer (4x): 225 mM Tris-HCl pH 6.8, 20% glycerol, 8% SDS, 0.04% bromophenol blue. 100 mM DTT was added at time of use.
- SDS Separating gel: 0.1% SDS, 0.05% ammonium persulfate, desired concentration of acrylamide and 375 mM Tris pH 8.8 and made up to 10 ml with distilled $H_2O$.
- SDS Stacking gel: 0.1% SDS, 0.05% ammonium persulfate, 4% Acrylamide, 125 mM Tris pH 6.8 and made up to 5 ml with distilled $H_2O$.
- Running buffer for western blotting: 25 mM Tris, 250 mM Glycine, 0.1% SDS.
- Transfer buffer for western blotting: 25 mM Tris, 250 mM Glycine.

**Tissue culture reagents and media**

- DMEM -Dulbecco's Modified Eagle Medium (Sigma) supplemented with 10% foetal bovine serum (Thermo Scientific) and Penicillin-Streptomycin-Glutamine (Gibco).
- Foetal bovine serum, heat inactivated (Sigma) used to supplement DMEM medium for OHT experiments.
- DPBS -Dulbecco's Phosphate Buffered Saline 1X without Ca or Mg (Corning).
- Trypsin-EDTA Solution 10X (Gibco) dissolved in DPBS to 1X working concentration.
- 4-hydroxytamoxifen, OHT (Sigma) was dissolved in DMSO and administrated at a final concentration of 600-780 nM.

- Geneticin, G 418 disulfate salt (Sigma) was administrated at a final concentration of 250 µg/ml.
- Puromycin (Thermo Scientific) was administrated at a final concentration of 2 µg/ml.
- Valproic acid (SIGMA) was dissolved in DPBS and administrated at a final concentration of 1 mM.
- Sodium butyrate (Sigma) was dissolved in DPBS and administrated at a final concentration of 3 mM.
- UNC 0638 (Sigma) was dissolved in DMSO and administrated at a final concentration of 250 nM.
- Hygromycin B (Thermo Scientific) was administrated at a final concentration of 200 µg/ml.

## 2.2 Methods

### 2.2.1 MEFs derived from mouse embryos and immortalization of cell lines

Mouse embryonic fibroblasts (MEFs) were derived from E13.5 stage embryos obtained from the crosses between heterozygous mice for the *Lsh*$^{off}$ or *Lsh*$^{on}$ allele and ER-Cre recombinase. The *Lsh*$^{on/on}$ MEFs were used for different experiments that were finally not included in this thesis. *Lsh*$^{+/+}$ MEFs and *Lsh*$^{off/off}$ MEFs were used for the experiments included in the coming chapters of the thesis.

Pregnant mice were killed by dislocation of the neck according to the manipulation instructions and the uterus with the embryos was collected with the help from Christian Belton. Embryos were dissected from the uterus without impairment of the amniotic membrane and the maternally derived blood tissue was washed 5 times with DPBS. The amniotic membrane was punctured and the embryos were placed in the individual dishes containing DPBS for further dissection. Then the tissues around the umbilical cord including placenta and foetal membranes were removed, the embryos were washed another 3 times with DPBS. These 13.5 days embryos were used to derive mouse embryonic fibroblasts (MEFs) by disruption of the body tissue with trypsin and subsequent culture with DMEM in tissue culture plates to select for MEFs.

Two different methods were used to immortalize these primary MEF cell lines: the ‘3T3 protocol' (Sharpless 2006) and retroviral transduction of cells with virus containing the SV40 Large T antigen.

### 2.2.2 Viral packaging and lentiviral transduction

Phoenix A cells were seeded and cultured for a few days in 10 cm tissue culture dishes before adding the plasmid. 15 µg of plasmid was transfected into the Phoenix A cells with the jetPEI DNA transfection reagent (PolyPlus). After 24 hours the medium was replaced and 2 days later the medium containing the lentiviral particles was collected and filter through a 0.4 µm filter (Minisart). The lentiviruses were aliquoted and frozen at -80°C.

Viral transduction of MEF cell lines was done in 5 ml of DMEM media (2.5 ml containing virus) with Polybrene at a final concentration of 5 µg/ml. Cells were incubated for 24 hours with the viral mixture, and then 5 ml extra of medium were added on top to avoid Polybrene toxicity and cells were again incubated overnight. Next day, antibiotic selection was started to select for stable clones expressing the plasmid.

### 2.2.3 Mammalian cell culture

All MEFs were cultured in Dulbecco's Modified Eagle Medium, DMEM (Sigma) containing 10% foetal bovine serum (Thermo Scientific) and Penicillin-Streptomycin-Glutamine (Gibco). Harvesting of the cells was done by trypsinization using 1 x Trypsin (Gibco). When treating MEFs in culture with tamoxifen, 4-hydroxytamoxifen (OHT) (Sigma) was dissolved in DMSO and administrated in a final concentration of 600-780 nM.

Phoenix A cells were cultured in the same conditions as MEFs and using the same reagents.

### 2.2.4 Molecular cloning

This method was used for generation of plasmids for transfection of cell lines, as well as for the bisulfite DNA sequencing. First, DNA was amplified by the Polymerase Chain Reaction (PCR), then restriction enzyme digestion was done when necessary, DNA product was run in an agarose gel and the desired bands were extracted to ligate the DNA into an appropriate vector. The final plasmids were normally sent for sequencing and purified before use by Miniprep or Maxiprep Kit (Quiagen) following the manufacturer’s instructions.

Polymerase chain reaction (PCR): DNA template was amplified in 1x buffer IV containing 12 mM $MgCl_2$, 2.5 mM dNTPs, 7% DMSO when needed, 5 µM primers and Taq polymerase (2.5 units for 50 µl reaction). The different steps of the chain reaction, denaturation, annealing and elongation; were carried out in a Biometra thermocycler and were dependent on the specific

reaction and application. For bisulfite sequencing, a special PCR buffer was used instead of buffer IV. The PCR products were analysed by agarose gel electrophoresis.

Restriction enzyme digestion: Restriction digests were carried out following the manufacturer's recommendations depending on the enzyme. Incubation times ranged from 1 hour to overnight at 37°C. Completion of digestion was assessed by agarose gel electrophoresis.

Gel extraction: The desired DNA bands were excised from agarose gels and DNA was extracted following the manufacturer's instructions using the GeneJet Gel Extraction Kit (Thermo Scientific).

Ligation: Purified DNA fragments were blunted using the DNA blunting enzyme (Thermo Scientific) and ligated into the desired linearized vector using the T4 DNA ligase (Thermo Scientific) with incubation times depending on experiment. For bisulfite sequencing the CloneJET PCR Cloning Kit (Thermo Scientific) was used. The ligation product was assessed by competent cell transformation and selection.

### 2.2.5 Genomic DNA extraction

Cell pellets were re-suspended in 1x TE buffer and Proteinase K (Sigma) and SDS were added to a final concentration of 200 µg/ml and 0.1%, respectively and incubated at 55°C overnight. Digested peptides were removed using phenol:chloroform:isoamyl alcohol extraction followed by chloroform extraction. The genomic DNA was precipitated with 1 volume isopropanol and 1/10 volume of 3 M NaOAc (pH 5.4) solution in a final concentration of 143 mM. Finally, the DNA pellet was washed by 70% ethanol and resuspended in desired volume of 1x TE buffer. RNA was removed by digestion with RNAse at 37°C for at least 30 minutes.

### 2.2.6 Extraction of nuclear proteins and protein quantification

Cell pellets were re-suspended in 1 ml Hypotonic NE1 buffer and disrupted by Dounce Homogeniser. The nuclei were pelleted by centrifugation at 4500 rpm for 5 minutes at 4°C and re-suspended in NE1 buffer adjusted with 5N NaCl to final NaCl concentration of 500 mM. Samples were incubated at $4^{o}$ C for 1.5 hours under rotation. Nuclear matrix and membranes were pelleted by centrifugation at 13000 rpm for 15 minutes at 4°C and supernatant containing the nuclear fraction was collected. Protein concentration was measured using the Bicinchoninic acid assay (BCA) method. Protein extracts were stored at -80°C.

### 2.2.7 Polymerase Chain Reaction –PCR

Polymerase chain reaction was performed in 1x normal buffer system containing $MgCl_2$, dNTPs, 7% DMSO when needed, Taq polymerase (2.5 units in 50 µl), 5 µM primers and specific amount of template with the conditions of denature, annealing and elongation depending on the specific amplification. The results of PCR were assessed by agarose gel electrophoresis.

### 2.2.8 Western Blotting

Nuclear extracts (55-60 micrograms of protein) were separated on 8% SDS-PAGE gel for LSH and 17% SDS-PAGE gel for histone modifications. The specific amount of protein was resuspended in a 1x sample buffer containing 0.1 M DTT and denatured at 95°C for 5 minutes. Different sizes of the protein were separated in a suitable concentration SDS-PAGE gel as mentioned above at 200 V for 60-75 minutes. Protein separated in the SDS gel was transferred onto Nitrocellulose or PDVF membrane (Bio-Rad) for 60-90 minutes in 1x cold transfer buffer at 400 mA. Membranes were blocked with 4% skimmed milk in 1x TBS and 0.1% Tween 20 for 1 hour at RT. Then the membranes were incubated overnight at 4°C in the freshly made blocking solution containing the desired concentration of primary antibody (see materials for antibody concentrations). After the overnight incubation, the membranes were washed three times with 1x PBS with 0.1% Tween 20 and then blocked again for 30 minutes. Secondary antibodies were added to a desired concentration (see materials for antibody concentrations) and the membrane was incubated for at least 1 hour. The membranes were washed as before to remove the unspecific binding of secondary antibodies and scanned on Odyssey 3.0 Scanner (LI-COR Biosciences).

### 2.2.9 RNA extraction

Cell pellets were re-suspended in 1 ml of Trizol reagent (Thermo Scientific) and 0.2 ml of chloroform. The samples were vigorously mixed and centrifuged at 12000 g for 15 minutes, 4 °C to extract the top aqueous phase that was precipitated with 0.5 ml of isopropanol. Then the precipitated samples were centrifuged at 12000 g for 10 minutes, 4 °C and RNA pellets were washed with cold 75% ethanol. Finally, the RNA pellets were dissolved in RNAse free water with 1 µl RiboLock RNAse inhibitor (Thermo Scientific). The trace amount of DNA was digested by 10 units of DNAse (Fermentas) at 37°C for 30-60 minutes and the DNAse was inactivated at 95°C for 3 min.

### 2.2.10 Quantitative Reverse Transcription PCR (qRT-PCR)

cDNA synthesis: 1 µg of oligo(dT) primer (0.5µM) was added to 4 µg of total RNA in a final volume of 12 µl. Samples were incubated at 65 °C for 5 min and cooled immediately on ice. 8 µl of mixture solution, including dNTPs (10 mM), 5x FS cDNA buffer, DTT (0.1 M) and RNAse inhibitor (40 U/ul), was added to each sample and incubated together at 42°C for 2 min. Finally, 1 µl (200 Units) of Superscript II/III Reverse Transcriptase (Invitrogen) was added selectively into the positive RT reactions and the reactions incubated at 42 °C for 120 min.

Quantitative PCR: qPCR was carried out on LightCycler 480 (Roche) with 1x SYBR Green I Master-mix (Roche) containing 2.5 µM primers and desired amount of template. The qPCR amplification for mRNAs of interest was calculated relative to GAPDH amplification as an internal control reference gene to determine differences in expression of the genes of interest.

### 2.2.11 Bisulfite DNA Sequencing

Bisulfite conversion of genomic DNA: 2 µg of genomic DNA was resuspended in TE buffer in a total volume of 25 µl. DNA was denatured at 110°C for 5 minutes and placed on ice. 2.5 µl of freshly made 3 M NaOH (final concentration of 278 mM) was added and the samples were incubated at 37°C for 20 minutes. Finally, 270 µl of sodium bisulfite solution (see solutions section) was added to each sample. The samples were overlaid with mineral oil and incubated overnight for sulfonation at 55°C. The treated genomic DNA was precipitated with 900 µl isopropanol and 3 M sodium acetate solution (final concentration 300 mM) with 5 µg glycogen. The DNA pellets were resuspended in TE buffer and desulfonated with freshly made 3 M NaOH for 15 minutes at 37°C. Samples were then precipitated using 32.5 µl of 5 M ammonium acetate pH 7.0 (final concentration 650 mM) and 180 µl 100% ethanol. Finally, DNA pellets were dissolved in 25 µl of TE buffer. The sodium bisulfite conversion of genomic DNA when sequencing *Rhox6/9* and *Gm9* loci was achieved using the EpiTect Bisulfite Kit (QIAGEN), following the manufacturers' instructions.

PCR amplification of regions of interest: Nested PCR primers were designed using Methprimer software aiming for amplification regions across CpG-rich region and PCR product length of around 300-400 base pairs long. 2 µl of sodium bisulfite treated DNA was used as template for the first round PCR reaction. For the nested PCR, 2-4 µl of the first round PCR product was used as template.

Gel extraction and PCR product purification: After running the PCR product in an agarose gel, the piece of gel containing the desired size of the DNA fragments was cut and melted at 55°C

for 10 min. The PCR product was purified according to the manufacturers' instructions using the Gene JET gel extraction kit (Thermo Scientific).

Ligation: The concentration of purified DNA fragments was measured using Pico-drop spectrophotometry. Ligation was performed according to manufacturer's instructions by using Clone JET PCR Cloning Kit (Thermo Scientific). DNA fragments were blunted and the pJET1.4 linearized plasmid and T4 DNA ligation enzyme were added. The ligation reaction was performed at 16°C for a minimum of 8 hours.

Transformation: 10 µl of ligation product was transformed into 50 µl of competent DH5α *E. coli* cells at 42°C for 60 seconds. After recovering at 37°C with gentle shaking for 1 hour, transformed *E. coli* were grown overnight on LB plates under Ampicillin selection.

Colony PCR: The forward and reverse primers for pJET1.4 plasmid were used for colony PCRs, and DNA from colonies of transformed *E. coli* was used as a template in the PCR reactions. For the colony PCR program an annealing temperature of 56°C was used.

Clean up of the colony PCR product: The PCR products were cleaned using a mixture containing 2 µl (40 units) of Exonuclease I (NEB), 20 µl (100 units) of Shrimp Alkaline Phosphatase (NEB) and 78 µl of $H_2O$. The mix was incubated at 37°C for 20 minutes and then at 80°C for 15 minutes.

DNA sequencing: Big Dye (Thermo Scientific) Reactions were set up in 10 µl including 4 µl of cleaned up colony PCR product. The following program was used for the reactions: 96°C for 1 minute and 25 cycles of cycling with each cycle at 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes. Finally, DNA sequencing was carried out by the Genepool Sequencing Service at the School of Biological Sciences, University of Edinburgh.

### 2.2.12 Chromatin-Immunoprecipitation (ChIP)

Cell pellets from a confluent T175 flask were resuspended in 9 ml of warm DPBS and 1 ml of 10x crosslinking buffer was added. The final concentration of formaldehyde was adjusted to 1% and the cell suspension was incubated for 5 minutes at room temperature. The crosslinking reaction was stopped by adding 550 µl of 2.5 M freshly made Glycine making a final concentration of 125 mM and the cells were collected by spinning at 1400 rpm for 6 minutes at 4°C and washed twice with cold DPBS containing 0.1 mM PMSF. The crosslinked cells were divided to $5 \times 10^6$ cells per tube aliquots and stored at -80°C until processed further.

Cells were thawed on ice and resuspended in 5 ml cold L1 buffer containing 0.1 mM PMSF (5 ml of buffer were used per $2 \times 10^6$ cells) and placed on a spinning wheel for 10 minutes at

4°C. Then, the cells were collected by spinning at 3000 rpm for 10 minutes at 4°C and washed with cold L2 buffer containing 0.1 mM PMSF and spun down as above. Finally, the cells were resuspended with 1850 µl of cold L3 buffer containing 0.5% SDS.

The cell suspension was sonicated by Biorupter device (Diagenode) at high frequency for 50 cycles with each cycle of 30 seconds on and 30 seconds off. 50-100µl of the sonicated chromatin was spun at 13000 rpm for 10 minutes and the supernatant was moved into a new tube. Equal volume of 1x TE buffer with 1% SDS and 5 µl of Proteinase K (Sigma) were added to decrosslink and digest the sample overnight at 65 °C. The rest of the sonicated chromatin was divided into equal aliquots (in order to have enough chromatin for a couple of IPs reactions) and stored at -80°C. After digestion, the DNA was extracted by the phenol/chloroform method (see methods for genomic DNA extraction) and the size of the sonicated chromatin was assessed by 2% agarose gel electrophoresis. For ChIP, the desired size of DNA fragments was around 300 base pairs.

Immunoprecipitation: The appropriate volume of sonicated chromatin samples, equivalent to 10 µg of DNA, was diluted in ChIP dilution buffer in order to reduce the SDS concentration to 0.1%. The volumes of the different samples were adjusted to become equal and appropriate antibody was added (4 µg / 500 µl IP reaction). Then the sonicated chromatin samples and the antibody were incubated for 2.5 hours on rotating wheel at 4°C. 50 µl of Protein G Dynabeads (Thermo Scientific) previously blocked with BSA (250 µg BSA for 100 µl beads) were added to each IP tube and incubated together for another 2.5 hours at 4°C. The beads were collected by a magnet and washed with Wash buffer 1 and 2 for 4 times and 3 times, respectively. The last wash was performed with 1 x TE buffer. The beads were then resuspended in 100 µl 1x TE buffer with 1% SDS final concentration and 5 µl of Proteinase K (Sigma, 20 mg/ml) was added. The beads were incubated at 65°C overnight. After digestion, the DNA fragments were extracted by PureLink Quick PCR Purification Kit (Thermo Scientific) and resuspended in 100 µl of distilled water.

Quantitative PCR (qPCR): the enrichment of DNA fragments corresponding to a genomic region of interest after ChIP was detected by qPCR. qPCR was carried out on LightCycler 480 (Roche) instrument with 1x SYBR Green I Master-mix (Roche) containing 5 µM primers and 2 µl of purified DNA fragments as template. The percent enrichment of regions of interest was calculated relative to input DNA amplification by comparative Ct method, also known as the 2–[delta] [delta] Ct method.

### 2.2.13 Micrococcal nuclease (MNase) digestion of chromatin

Collection of cells: 5 × $10^6$ cells were harvested, crosslinked as for ChIP and washed twice with ice-cold 1x DPBS and spun down. At this point, the cell pellets were be kept at -80°C if not processed immediately.

MNase treatment: the cells were resuspended in 500 µl of douncing buffer (10 mM Tris-HCl pH 7.5, 4 mM $MgCl_2$, 1 mM $CaCl_2$, and 1x Protease Inhibitors Cocktail from Sigma added fresh) and homogenized in a Dounce homogeneizer with 25 strokes. Next, 200 Units of MNase (Thermo Scientific) were added and the samples were incubated at 37°C for 15 minutes while shaking at 800 rpm. After the incubation, the reaction was quenched by adding EDTA to a final concentration of 5 mM and incubation on ice for 5 minutes. Samples were digested overnight with 10 µl Proteinase K and 1% SDS at 65°C. Next day, DNA was purified by Phenol/Chloroform extraction and incubated with 0.1 µg/µl RNaseA at 37°C for 30 minutes.

Assessment of MNase digestion: DNA samples were run on 1.5% agarose gel to determine the size of nucleosomal DNA bands. Mononucleosomal fragments (100-200 base pairs) were excised and DNA purified with GenJet Gel Extraction Kit (Thermo Scientific) when needed.

### 2.2.14 Nucleosome Occupancy and DNA Methylation (NOMe)

Collection of cells and crosslinking: exponentially growing cells were harvested, crosslinked following the ChIP protocol and washed twice with ice-cold 1x DPBS. Then, cells were spun down and resuspended in 1 ml of ice-cold Nuclei Buffer (100 mM Tris-HCl pH 7.4, 100 mM NaCl, 30 mM $MgCl_2$, and 1x Protease Inhibitors Cocktail from Sigma added fresh).

Nuclei Extraction: samples were incubated on ice for 5 min minutes and homogenized with 15 strokes of a Dounce homogenizer. The homogenized cells were transferred into a tube and centrifuged to recover nuclei at 3500 rpm for 3 minutes at 4°C. Cells were washed in Nuclei Buffer and resuspended in 1x M.CviPI reaction buffer (NEB) so that there were $10^6$ cells per 39.25µl.

Treatment with M.CviPI: The nuclei previously resuspended was incubated for 15 minutes at 37°C with 200 Units of M.CviPI enzyme in 1x M.CviPI Buffer containing 160 µM SAM, 300 mM Sucrose, and water to achieve 150 µl final volume for the reaction. The reaction was stopped by adding an equal volume (150 µl) of Stop Solution (20 mM Tris-HCl pH 7.9, 600 mM NaCl, 1% SDS, 10 mM disodium EDTA, and Protease Inhibitors Cocktail (Sigma) added fresh) and the decrosslinking was done overnight incubating at 65°C with Proteinase K (Sigma).

Purification of DNA by Phenol/Chloroform extraction and ethanol precipitation was done next, and purified DNA kept at -20°C up to 1 year if not used.

Bisulfite DNA Conversion was done with the EpiTect Bisulfite Kit (Qiagen) and PCR for specific region / promoters of interest as well as cloning for sequencing were performed following the Bisulfite Sequencing protocol described above.

# Chapter 3 - LSH can silence active promoters

## 3.1 Introduction

Previous work in the Stancheva laboratory demonstrated that LSH is required for DNA methylation at centromeric repeats, repetitive elements and some specific gene promoters (Myant et al., 2011). It was previously thought that in the absence of LSH, defects in DNA methylation and transcriptional expression were limited to repetitive elements and a few single copy sequences (Dennis et al., 2001). In addition, the study carried out in the laboratory showed that LSH is required for DNA methylation and appropriate gene expression not only at repetitive elements such as centromeric and pericentric satellite repeats, IAP retrotransposons, and LINE elements but also at protein coding genes. More recent studies have also shown that specific protein coding genes are affected by lack of LSH as well as repetitive sequences and large chromosomal domains associated with the nuclear lamina (Yu et al., 2014). Furthermore, the study carried out in the lab also compared CpG DNA methylation at promoters of protein coding genes in wild-type and *Lsh*$^{-/-}$ MEFs. It was found that 20% of normally methylated promoters had loss of DNA methylation in the absence of LSH (Myant et al., 2011). Besides the loss of DNA methylation, many of these genes were aberrantly expressed in the *Lsh* knockout cells. This correlates with the gain of transcription observed at repetitive sequences in the absence of LSH. Some of these specific genes affected by LSH are the *Rhox* genes that are part of the reproductive homeobox X-linked gene cluster *Rhox*, *Gm9* adjacent to the *Rhox* cluster, and development and pluripotency associated genes such as *Dppa2* and *Dppa4*. The Rhox cluster and the neighbouring *Gm9* gene, show a big reduction in DNA methylation in *Lsh*$^{-/-}$ MEFs, being the largest contiguous genomic region affected by the lack of LSH in these cells.

The Rhox cluster encodes a group of reproductive homeobox genes on the X chromosome that are expressed in a cell type specific manner. Most of these genes are only expressed in germ lines, in both male and female mice (MacLean et al., 2005). The promoters of *Rhox* genes are hypomethylated in the pre-implantation embryo and acquire DNA methylation after implantation by the E9.5 embryo stage. This gain of DNA methylation happens in the embryonic but not in the extraembryonic tissues (Oda et al., 2006). The *de novo* DNA methylation establishment at the promoters of *Rhox* genes is dependent on DNMT3B and partly on DNMT3A, but is independent of X chromosome inactivation since the *Rhox* genes are methylated and silenced in both male and female somatic cells. The DNA methylation status of these genes distinguishes early embryonic lineages (Oda et al., 2006),

suggesting that LSH is implicated in *de novo* DNA methylation and gene expression regulation during early development.

Previous work performed in the lab (Termanis et al., 2016) used wild-type (WT) and *Lsh* $^{-/-}$ MEFs that were provided by Kathrin Muegge lab. *Lsh* knockout mice were obtained crossing heterozygous mice for the knockout allele, meaning that wild-type LSH was present during germ cell determination in the parents. As a consequence, the DNA methylation abnormalities observed in the *Lsh*$^{-/-}$ MEFs can only be attributed to defects in the maintenance of DNA methylation patterns established during maturation of germ cells or *de novo* methylation defects at later stages of embryo development. It is also important to mention that *Lsh*$^{-/-}$ embryos survive until birth without major developmental defects. This might indicate that the incomplete DNA methylation establishment in the absence of LSH do not cause substantial developmental defects during embryonic development.

Given the existing knowledge, this study aimed to address two major questions. First, whether LSH has an activity in somatic cells where the signalling and transcription factors characteristic of early development are largely absent. Second, whether the ATP hydrolysis, i.e. the catalytic activity of LSH, is necessary for the role of LSH in DNA methylation and gene silencing. In order to answer these questions, stable cell lines expressing different LSH plasmids were generated to better understand and potentially rescue the defects in *Lsh*$^{-/-}$ MEFs. Two plasmids, pMSCV-LSH and pMSCV-K237Q were generated by cloning into pMSCV vectors, and both plasmids were stably expressed in *Lsh*$^{-/-}$ MEFs. The pMSCV-LSH plasmid contains wildtype LSH. The pMSCV-K237Q plasmid contains a mutation in the catalytic site converting a conserved lysine into glutamine in the catalytic domain of LSH making this mutant unable to bind and hydrolyse ATP. An additional control cell line was generated with the empty pMSCV vector and used as a control knockout cell line, *Lsh*$^{-/-}$ MSCV MEFs. I will refer to this cell line in the text as *Lsh*$^{-/-}$ MEFs. Experiments using these rescue cell lines showed that the gain of expression at specific protein coding genes and repetitive elements occurring in the absence of LSH can be rescued to the initial silenced state when wild-type LSH is expressed in *Lsh*$^{-/-}$ MEFs but not the catalytically inactive LSH (K237Q). Among these genes are the reproductive homeobox (*Rhox*) genes and the *Gm9*. Importantly, this gene silencing occurred in somatic cells, *Lsh*$^{-/-}$ LSH MEFs, in the absence of differentiation signals, suggesting that LSH plays a role in a cell autonomous manner. The promoters of these unique genes and repetitive elements misregulated in the absence of LSH are characterised by overall 40-50% loss of DNA methylation at their promoter region. Accordingly, when wild-type LSH was introduced in *Lsh*$^{-/-}$ MEFs, there was transcriptional silencing of these loci and the *de novo* DNA methylation was re-established at these promoters. However, the catalytically inactive LSH was not able to rescue DNA methylation,

indicating that ATP hydrolysis by LSH is necessary for both gene silencing and establishment of *de novo* DNA methylation in *Lsh*$^{-/-}$ MEFs. Bisulfite sequencing analyses for IAPs detected almost complete methylation at promoters in some of the sequenced clones and high gain of DNA methylation, although not complete, at the promoters of the *Rhox* genes that were silenced such as *Rhox2, Rhox6* and *Rhox9* genes. On the other hand, some pluripotency associated genes such as *Dppa2* and *Dppa4* were not transcriptionally silenced in *Lsh*$^{-/-}$ LSH MEFs. This indicates that additional signals, not present in MEFs, might be required for silencing of pluripotency associated genes in somatic cells.

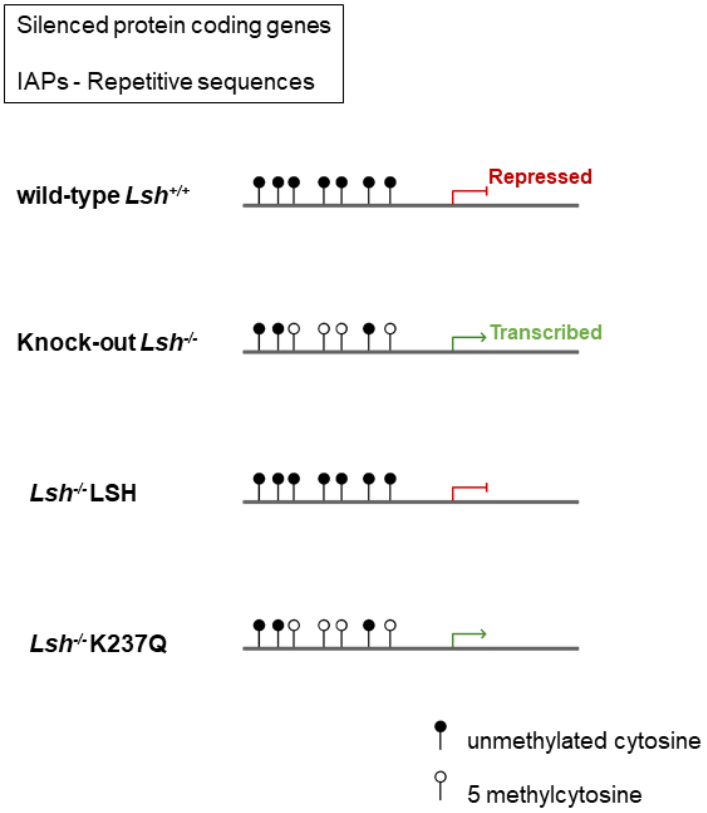


**Figure 3.1. Schematic diagram showing LSH-dependent gene silencing of specific protein coding genes and IAP retrotransposons.**

Representation of bisulfite DNA sequencing in wild-type MEFs, *Lsh* knockout MEFs, and wild-type LSH or mutant LSH rescue cell lines. Each row indicates a DNA strand representing the promoter region of LSH-dependent loci and each circle is a CpG dinucleotide. Methylated CpGs are shown in black and unmethylated CpGs in white. LSH is required for appropriate *de novo* DNA methylation at these promoters and gene silencing. Wild-type LSH is able to rescue the aberrant DNA methylation and active transcriptional state of knockout cells. Mutant LSH (K237Q) is not able to rescue the wild-type pattern in these cells.

In summary, previous work in the lab has characterised LSH-dependent loci including not only repetitive elements that had been already characterised in many studies before but also protein coding genes. Extensive DNA methylation and expression defects in *Lsh*$^{-/-}$ MEFs were characterised genome wide. This shows that LSH is directly involved in regulating DNA methylation and gene expression throughout the genome at single gene promoters as well as clusters of promoters and repetitive sequences (Myant et al., 2011). LSH can overcome these defects cell autonomously, when introduced into the *Lsh*$^{-/-}$ MEFs. However, the catalytic activity of LSH is necessary for this restoration of DNA methylation and transcriptional silent state (Figure 3.1) (Termanis et al., 2016). It is still unknown how LSH enables this gene silencing and DNA methylation to occur at promoters of some genes such as the *Rhox* cluster in somatic cells but not others such as the pluripotency associated genes. It would also be interesting to determine whether histone modifications are involved in LSH-dependent gene silencing and whether there is a difference in these marks between the loci that respond differently to restoration of LSH expression. Answering these questions will be the aim of this chapter. The experiments described in this chapter have been published in Nucleic Acids Research (Termanis et al., 2016), see appendix 1.

### 3.2 LSH enables silencing of active promoters marked by H3Ac and H3K4me3

To better understand how LSH contributed to gene silencing at LSH-dependent loci, it was important to analyse whether active transcription of these specific genes in the *Lsh*$^{-/-}$ MEFs was accompanied by histone modifications associated with this transcriptional state. In order to determine whether these loci were marked in their promoter region by active histone marks such as histone H3 lysine 4 trimethylation (H3K4me3) and histone H3 lysine 9 / lysine14 acetylation (H3Ac), I carried out ChIP for both marks. Both histone marks, H3K4me3 and H3Ac, are normally found in promoters of active genes. H3K4me3 is also found as a mark in unmethylated CpG islands in mammalian cells, and this modification does not necessarily correlate with its transcriptional activity (Thomson et al., 2010). ChIP experiments were done using chromatin extracted from crosslinked cells collected from wild-type MEFs as well as *Lsh*$^{-/-}$ MEFs and both rescued cell lines, *Lsh*$^{-/-}$ LSH MEFs and the catalytically inactive *Lsh*$^{-/-}$ K237Q MEFs. These ChIP experiments showed that *Lsh*$^{-/-}$ MEFs were marked by both active histone modifications compared to wild-type MEFs where we did not observe these active marks (Figure 3.2). However, the enrichment for these marks in *Lsh*$^{-/-}$ MEFs was different depending on the specific locus analysed. The higher enrichment for both active histone marks at some genes did not seem to be accompanied by higher transcription (Termanis et al., 2016). The transcriptional silencing and DNA methylation at the promoters of *Gm9* and *Rhox* genes in

*Lsh*$^{-/-}$ LSH MEFs was accompanied by loss of both active marks, H3K4me3 and H3Ac marks, with the exception of *Rhox2* which does not show a rescue in these marks (Figure 3.2). The misregulated genes, which were not silenced in the *Lsh*$^{-/-}$ K237Q MEFs, do not lose H3K4me3 and H3Ac marks from their promoters. In contrast, these *Lsh*$^{-/-}$ K237Q MEFs showed an increase in both histone marks at the majority of promoters which matched the increase in gene expression, although some genes such as *Rhox2* and *Rhox9* do not show this increase. The lack of increase in both histone marks at *Rhox2* and *Rhox9* in the *Lsh*$^{-/-}$ K237Q MEFs suggests that methylation and histone modifications are not coupled at these loci. There might be different models of action on different genes.

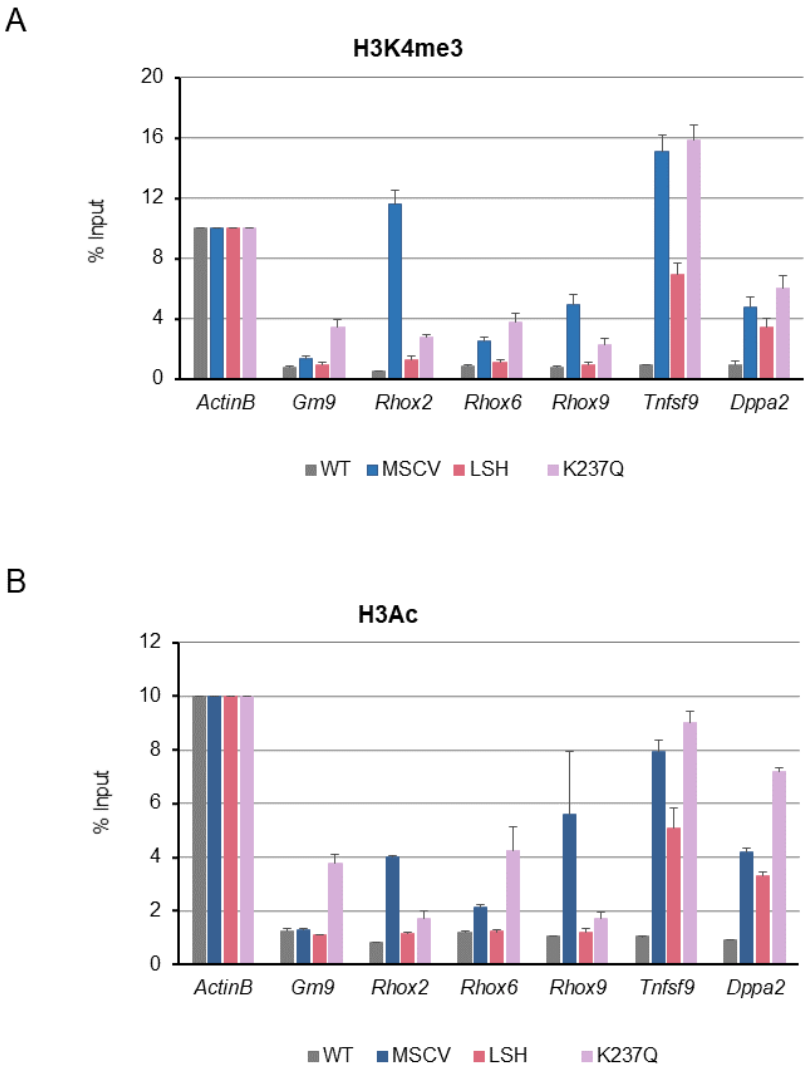


**Figure 3.2. LSH enables silencing of active promoters marked by H3K4me3 and H3Ac.**

The silencing of gene expression by wild-type LSH is accompanied by loss of both marks, H3K4me3 and H3Ac, from gene promoters. (A) ChIP for H3K4me3. (B) ChIP for H3Ac. *Actin* was used as a control since its transcriptional state does not change in the cells in study. *Tnfsf9* was randomly chosen from a list of promoters that acquire DNA methylation in the wild-type LSH rescue cells. A rabbit IgG antibody was used as an additional control for H3K4me3 antibody and H3Ac antibody (data not sown). The error bars represent standard deviation calculated from 2 biological replicas with 2 technical replicates each.

As mentioned earlier, pluripotency-associated genes such as *Dppa2* and *Dppa4* did not follow the same pattern as *Rhox* genes and were not silenced in *Lsh*$^{-/-}$ LSH MEFs. Accordingly, the ChIP experiments showed that the active histone marks at the promoter region of *Dppa2* did not show a big decrease compared to the *Rhox* genes or *Gm9* when LSH was re-introduced. However *Dppa2* did show an increase in both H3K4me3 and H3Ac in the *Lsh*$^{-/-}$

K237Q MEFs, similar to the increase observed at the *Rhox* gene promoters (Figure 3.2). I also analysed two additional loci, *Actin* and *Tnfsf9*, which were used as controls for these ChIP experiments. *Actin* is expressed in both wild-type MEFs and *Lsh*$^{-/-}$ MEFs (Myant et al., 2011) and its expression did not change in any of the rescued cell lines, *Lsh*$^{-/-}$ LSH MEFs or *Lsh*$^{-/-}$ K237Q MEFs. *Tnfsf9* was randomly chosen as an additional control since it behaves in the same way as *Dppa2. Tnfsf9* has no DNA methylation at the promoter and it is expressed in the absence of LSH. However, when LSH is reintroduced in the knockout cells, DNA methylation is not rescue at *Tnfsf9* promoter. We are not aware of its expression state in these cells but it seems that *Tnfsf9* follows the same trend as pluripotency genes, since we can still observe the active marks in *Lsh*$^{-/-}$ LSH MEFs.

Taken together, from these ChIP experiments we concluded that LSH when re-expressed in LSH-deficient cells is able to silence aberrantly expressed genes that are marked at their promoter region by active histone modifications H3K4me3 and H3Ac. These marks decrease only when wild-type LSH, but not the catalytically inactive mutant LSH, is expressed in *Lsh*$^{-/-}$ MEFs although there is a variable response of some genes such as *Rhox2*. The mutant LSH does not cause silencing and most of the loci do not lose active chromatin marks from their promoters demonstrating that the ability of LSH to hydrolyse ATP is required for these events.

### 3.3 *Lsh*$^{-/-}$ MEFs expressing wild-type LSH require additional signals for *de novo* DNA methylation and silencing of pluripotency associated genes

As mentioned earlier in this chapter, some loci, in contrast to the *Rhox* genes, do not gain silencing when wild-type LSH is re-expressed in the *Lsh*$^{-/-}$ cells. Following this, it was interesting to know whether the *de novo* DNA methylation was established or not in the promoters of *Dppa2* and *Dppa4* even though the genes were not silenced. In order to study this, I designed primers for bisulfite sequencing for both loci and optimized the technique. After doing the bisulfite conversion and DNA sequencing using gDNA from wild-type, *Lsh*$^{-/-}$ MEFs and both rescued cell lines, LSH and K237Q; I found that there was no rescue of DNA methylation in the promoter region of *Dppa2* and *Dppa4* (Figure 3.3 A / B). In the *Dppa2* gene promoter region we observed a decrease in DNA methylation by 59% in *Lsh*$^{-/-}$ cells compared to wild-type MEFs. The percentage of DNA methylation in *Lsh*$^{-/-}$ LSH and *Lsh*$^{-/-}$ K237Q MEFs was 17% and 18%, respectively, being similar to 21.7% DNA methylation in *Lsh*$^{-/-}$ cells (Figure 3.3 A). In the *Dppa4* gene promoter region the loss of DNA methylation was higher, observing a decrease by 80% in DNA methylation in *Lsh*$^{-/-}$ cells compared to wildtype MEFs. In a similar way to what happens at *Dppa2*, DNA methylation was not established back at *Dppa4* promoter in either *Lsh*$^{-/-}$ LSH or *Lsh*$^{-/-}$ K237Q MEFs, being the percentages of methylation 25% and 9%

respectively (Figure 3.3 B). This indicates that *Dppa2* and *Dppa4* might need additional signals that are not present in somatic cells to both, silence and re-establish *de novo* DNA methylation. In order to test this hypothesis, we treated *Lsh*$^{-/-}$ LSH and *Lsh*$^{-/-}$ K237Q MEFs with retinoic acid, since previous studies had shown that retinoic acid regulates pluripotency genes (Maldonado-Saldivia et al., 2007). Particularly, the expression of *Dppa5* has been shown to be downregulated during induced differentiation of ES cells with retinoic acid (Kim et al., 2005).

*Lsh*$^{-/-}$ MEFs expressing either wild-type or mutant LSH were treated with 300 nM retinoic acid for 5 days and this treatment led to silencing of *Dppa4* and *Dppa2* (data only shown for *Dppa4*) . However, the silencing was stably maintained only by cells expressing the wild-type LSH but it could not be maintained by cells expressing the K237Q mutant LSH (Figure 3.3 C). 5 days after removal of the retinoic acid, *Dppa4* expression gradually increased in the *Lsh*$^{-/-}$ K237Q cells and 10 later, the expression of *Dppa4* in *Lsh*$^{-/-}$ K237Q cells reached similar levels of expression as in the non-treated cells. In contrast, the silencing was stably maintained even 10 days after removal of retinoic acid in the *Lsh*$^{-/-}$ LSH MEFs. We asked whether the gene silencing observed when treating these cells with retinoic acid and the difference observed between LSH and K237Q rescued cell lines when removing the treatment, were also accompanied by a difference in the establishment of the *de novo* DNA methylation at *Dppa4* promoter. Analysis of DNA methylation by bisulfite DNA sequencing of *Dppa4* promoter carried out 10 days after removal of retinoic acid from the cell culture, showed that DNA methylation was restored only in the *Lsh*$^{-/-}$ LSH cells (74%) to similar levels as in the wild-type MEFs (75.8%) (Figure 3.3 D). However, DNA methylation in the *Lsh*$^{-/-}$ K237Q MEFs was 9.8%, showing no change in comparison with the methylation levels before the retinoic acid treatment. These results demonstrate that the ATPase activity of LSH is required for both gene silencing and the establishment of *de novo* DNA methylation. Importantly, DNA methylation at the promoter region is necessary for the long-term stable maintenance of silenced state of genes.

In conclusion, pluripotency associated genes in *Lsh*$^{-/-}$ MEFs require additional signalling, when the expression of LSH is restored, to rescue their silent state. Also, the ability of LSH to hydrolyse ATP is necessary for this silencing and for the establishment of *de novo* DNA methylation at the promoter region of *Dppa4*.

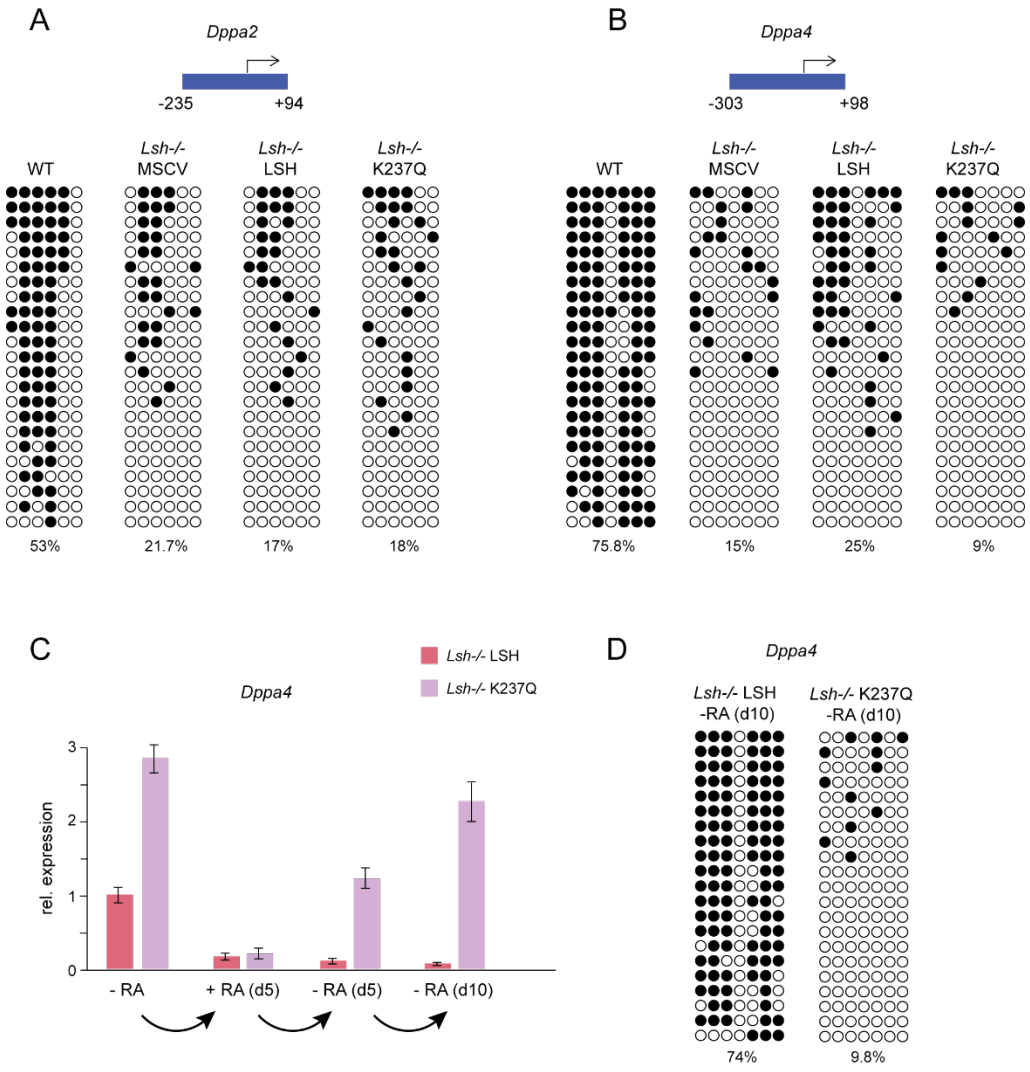


**Figure 3.3. Lsh knockout MEFs expressing wild-type LSH require additional signals for silencing of pluripotency associated genes.**

Bisulfite DNA sequencing of *Dppa2* (A) and *Dppa4* (B) promoters in wild-type MEFs, *Lsh* knockout MEFs carrying an empty MSCV vector and *Lsh* knockout MEFs expressing either wild-type or mutant LSH. Each row of circles represents a single DNA strand, methylated CpGs are shown in black and unmethylated CpGs are shown in white. The percentage of methylated CpGs is shown below the diagrams. The numbers below the promoter graphs indicate bp upstream (-) or bp downstream (+) from the transcription start site. (C) Treatment of wild-type or mutant LSH rescued MEFs with retinoic acid. The graph shows the expression of *Dppa4* relative to *Gapdh* measured by qRT-PCR. The error bars represent standard deviation calculated from 2 biological replicas with 3 technical replicates carried out for each time point. (D) Bisulfite DNA sequencing of the promoter of *Dppa4* analysed 10 days after removal of retinoic acid from the culture. Percentage of methylation is shown below the diagrams.

### 3.4 Discussion

In summary, previous work carried out in the lab has shown that specific genes misregulated in the *Lsh*$^{-/-}$ MEFs can be silenced and DNA methylation restored at their promoters when the wild-type LSH but not the catalytically inactive LSH K237Q is introduced into these cells. We were surprised to obtain this result since previous studies had been done in differentiated ES cells but never in somatic cells. We were asking a different question, whether LSH reintroduced in *Lsh*$^{-/-}$ somatic cells could rescue the inappropriate expression of LSH-dependent genes in a cell autonomous manner. We have shown that LSH is able to rescue both the appropriate expression and DNA methylation in a cell autonomous manner, but the catalytic activity of LSH is required for it. Accordingly, other study showed that chromatin remodelling by LSH was necessary for its function in DNA methylation (Ren et al., 2015). From the ChIP experiments described in this chapter we found that that the aberrant expression of specific developmentally-regulated genes in *Lsh*$^{-/-}$ cells is also accompanied by a gain in active histone marks H3K4me3 and H3Ac. This is in agreement with previous observations that H3K4me3 was increased in genes showing upregulation in *Lsh* knockout cells but not in genes downregulated in the same condition (Tao et al., 2011). Surprisingly, LSH is able to silence genes that are marked in their promoter region by active histone modifications such as H3K4me3 and H3Ac. Moreover, these modifications decrease in the specific genes where restoration of LSH expression promotes gene silencing but not in the control gene *Actin* or *Dppa2* which are not silenced when wild-type LSH is reintroduced into the *Lsh*$^{-/-}$ cells. It would be interesting to study the expression of *Tnfsf9* in these rescue cell lines since the histone modifications found at its promoter behave in a similar way as *Dppa4.* This might indicate that *Tnfsf9* does not get silenced when wild-type LSH is reintroduced in *Lsh*$^{-/-}$ MEFs as we observed for the development and pluripotency associated genes. To summarise, the ChIP experiments show that silencing of the genes in study is accompanied by loss of active chromatin marks. In relation to this, it is also important to mention that the promoter of *Rhox* genes has a high number of CpGs, being closer to a CpG island content. CpG islands are generally associated with promoters and have a GC percentage greater than 50%, and an observed-to-expected CpG ratio greater than 60% (Saxonov, Berg, & Brutlag, 2006). However, the number of CpGs at the promoters of the pluripotency associated genes *Dppa2 and Dppa4* analysed here is very low. This could explain why the enrichment for H3K4me3 is higher for some *Rhox* genes in *Lsh*$^{-/-}$ cells since H3K4me3 can be found as a mark of CpG islands independently of its transcriptional state (Clouaire et al., 2012; Thomson et al., 2010). Moreover, it is not clear why both active histone marks increase in the catalytically inactive LSH rescued cell line. From previous studies it is known that the ATPase activity of LSH is not required for its recruitment to chromatin, but it is important for the release of the enzyme

from these sites (Lungu, Muegge, Jeltsch, & Jurkowska, 2015). The ATPase-deficient mutant has a stronger association with heterochromatin. This stronger association could explain why there is no silencing and why gene expression and active marks are higher in *Lsh* $^{-/-}$ K237Q MEFs. The catalytically inactive LSH might still be recruited to these loci in the same way as wild-type LSH and more stably associate at these locations of the genome for a longer period. This could prevent the access of other enzymes or repressive complexes that might be acting in the knockout cells to control the altered gene expression. Another possibility is that the mutant LSH, which binds to DNA, might still bind HDACs or DNMT3B through its N-terminal domains in the same way as wild-type LSH does. This should be the case since the N-terminal region of LSH is sufficient for its interaction with HDACs and DNMT3B. However, the interaction of the mutant LSH with these proteins has to be proven since it has never been tested before. If this is the case, the mutant LSH could be impairing the activity of these enzymes by capturing them together with its inability to release from chromatin. Further experiments should be performed in order to clarify whether mutant LSH can still bind HDACs and DNMT3B, and to shed more light on the enhanced expression of LSH-dependent genes in *Lsh*$^{-/-}$ cells expressing catalytically inactive LSH.

It has been established that embryonic development is a highly orchestrated process where signalling pathways and transcription factors play an important role in defining the gene expression pattern in differentiating cells (Boyer et al., 2006; Loh et al., 2006). We have shown that LSH is able to re-establish proper DNA methylation and gene expression patterns at the *Rhox* cluster and *Gm9* gene in somatic cells lacking the specific signalling and transcription factors characteristic of embryonic development. However, pluripotency associated genes required additional signalling for establishment of DNA methylation and gene silencing as I showed by qRT-PCR analyses and bisulfite DNA sequencing after retinoic acid treatment. Silencing of pluripotency genes during development and differentiation of stem cells involves transcriptional regulators and various signalling pathways (Boiani & Schöler, 2005). LSH is expressed at high levels in ES cells, before silencing of pluripotency genes occurs. This suggests that initiation of the inactivation of pluripotency genes could require exogenous signalling factors characteristic of ES cells. These developmental specific signalling pathways of stem cells could be a requirement for LSH-dependent gene silencing of specific genes. However, exogenous signalling molecules are absent in MEFs. Previous studies have found that retinoic acid signalling has a complex function during vertebrate development and together with other factors facilitates reprogramming in MEFs (W. Wang et al., 2011). Controlled low levels of retinoic acid can modulate Wnt signalling and are required for the reprogramming of epiblast stem cells into embryonic stem cell-like cells, highlighting the important functions of retinoic acid signalling in reprogramming somatic cells (Yang et al.,

2015). In summary, silencing of pluripotency associated genes triggered by retinoic acid in wild-type LSH rescued cells, indicates that exogenous signalling characteristic of embryo development is required for LSH-dependent silencing of some loci. In addition, it reinforces the idea mentioned in the introduction of this chapter about the importance of LSH during early embryo development as previously reported for the silencing of *Hox* genes (Xi et al., 2007).

Another aspect which is not yet understood is how LSH is recruited to these specific loci in the genome. Unlike other SNF2-like family members involved in transcriptional regulation, LSH does not have a characterized chromatin binding domain such as bromodomain or chromodomain (T M Geiman et al., 1998). It should be noted that in the knockout cells, the expression of wild-type LSH did not recue all misregulated genes. A possible explanation for this is impairment of the recruitment of LSH to some areas of the genome. However, retinoic acid treatment triggered the establishment of gene silencing and *de novo* DNA methylation at pluripotency genes. This indicates that the lack of appropriate exogenous signalling in somatic cells could be an explanation rather than inaccessibility of specific areas of the genome to chromatin remodelling by LSH. It is not clear how LSH is recruited to the specific genes where we observe an effect. One hypothesis could be that LSH is recruited to chromatin by DNMTs which bind to chromatin through the ADD domain. The ADD domain in the *de novo* DNMTs preferentially binds histone peptides containing unmodified histone H3 lysine 4, H3K4me0 (Otani et al., 2009). We have shown that appropriate concentration of DNMT3B is necessary for rescue of silencing and methylation by wild-type LSH (Termanis et al., 2016). LSH interacts through its N-terminal region (coiled-coil domain) with DNMT3B as indicated in a previous study (Myant & Stancheva, 2008). DNMTs, specifically *de novo* DNMTs, have high affinity for chromatin (Noh et al., 2015) in contrast to LSH that has a lower affinity for chromatin and displays a very dynamic behaviour (Lungu et al., 2015). One possible explanation for the recruitment of LSH to specific areas of the genome could be the binding patterns of DNMTs in the *Lsh*$^{-/-}$ MEFs. If DNMTs occupancy in these cells remains unchanged, this will direct LSH to the areas where it is needed thus enabling re-establishment of DNA methylation. However, the presence of active marks at some of the LSH-targeted genes in *Lsh*$^{-/-}$ MEFs rules out this hypothesis since the binding of DNMT3A and DNMT3B to the N-terminal tails of histone H3 is inhibited by the methylation of lysine 4 , both H3K4me2 and H3K4me3 (Otani et al., 2009). Further experiments should be performed to address how LSH is recruited to these specific loci and understand how LSH is mediating these changes of chromatin modifications in somatic cells.

It is still unknown whether the observed changes in active histone marks are a secondary effect of the gene silencing caused by re-establishment of DNA methylation or a direct effect

of LSH being restored in these cells. Previous work in the lab, has shown that in reporter-based assays the interaction of LSH with histone deacetylates (HDACs) is necessary for LSH-dependent transcriptional repression (Myant & Stancheva, 2008) that does not involve *de novo* DNA methylation. Therefore further studies should be carried out to investigate the order of events leading to LSH-dependent gene silencing at endogenous genomic loci since the mechanism for this epigenetic repression is yet unclear. It has been suggested in many publications that the main role of LSH in epigenetic silencing is to facilitate DNA methylation by DNMT3B, but other studies have also shown that the LSH-dependent repression of retrotransposons, such as IAPs, does not require the function of DNMT3B (Dunican et al., 2013). Future work should focus on genome-wide approaches aiming to characterise what happens at LSH-dependent loci when LSH is re-expressed. We know that on specific protein coding genes, which are affected by the absence of LSH, DNA methylation and gene expression can be rescued in a cell autonomous manner when LSH is re-expressed. Genome-wide analysis would help clarify at which LSH-dependent loci this rescue happens. Once these loci that are able to re-establish normal patterns of expression and DNA methylation in LSH-recued cells are characterised, ChIP-Seq for LSH should be performed to clarify whether it binds preferentially to specific regions of the genome. These approaches will help understanding the dynamics of LSH-dependent transcriptional silencing in a genome-wide context. Also, this study focuses on *Lsh*$^{-/-}$ MEFs but in order to confirm a potential role of LSH in DNA methylation and gene silencing during early development, studies in embryos and analyses during different developmental stages should be performed. The use of a conditional knockout system that allows the study of embryos with the potential to reactivate LSH at different developmental stages would provide a better characterization of the developmental timing of LSH-dependent events. However, embryo studies can be difficult to carry out while conditional knockout cell lines could provide insight into the order of LSH-dependent events leading to gene silencing and *de novo* DNA methylation. A conditional knockout system for LSH needs to be established, validated and characterised before performing the analyses suggested above. The establishment of such an *in vitro Lsh* conditional knockout system will be the focus of chapter 4.

# Chapter 4 - Establishment of an *in vitro* conditionally reversible *Lsh* knockout system

## 4.1 Introduction

The previous chapter raised the need to establish a conditional reversible *Lsh* knockout system in order to better understand the role of LSH in epigenetic gene silencing. Such system would help to gain insight into the specific time at which LSH is required during embryo development and the order of events leading to LSH-dependent gene silencing and gain of *de novo* DNA methylation. In order to study these aspects and clarify the role of LSH in epigenetic gene silencing, a cell system established from mice carrying conditionally reversible *Lsh* alleles and tamoxifen-inducible Cre recombinase had to be optimised.

Chao Li, a former postdoc in the Stancheva lab, designed and generated a cassette flanked by different recombinase target sites (RTs) based on an approach previously used by Schnütgen et al (Schnütgen et al., 2005). The cassette contains a splice acceptor (SA), green fluorescent protein gene (GFP), neomycin resistance gene (neo), and polyadenylation (pA) sequence. I will refer to this cassette as the GFPneo cassette. The design of this cassette was based on previously reported gene trap vectors for conditional mutagenesis in embryonic stem (ES) cells. These vectors rely on directional site-specific recombination systems that can repair and re-induce gene mutations when activated in succession. Once the vectors are inserted into the mouse genome, genetic mutations can be produced at a particular time and place in somatic cells. The recombinase target sites flanking the cassette designed in the lab were the following: frt and F3 (heterotypic target sequences for the FLP recombinase), loxP and lox511 (heterotypic target sequences for the Cre recombinase). This cassette was introduced by homologous recombination into the intron 3 of the *Lsh* gene, generating heterozygous targeted ES cells from which a strain of mice carrying conditionally reversible $Lsh^{off}$ allele was established. The $Lsh^{off/+}$ mice were then crossed with a strain carrying ER-Cre transgene to produce $Lsh^{off/+}$, ER-$Cre^{+}$ mice. The *Lsh* gene is conditionally inactivated by the SA and pA in the cassette ($Lsh^{off}$) since transcripts initiated at the endogenous promoter are spliced from the splice donor (SD) of the endogenous exon 3 to the SA of the cassette. Then the GFP and neomycin resistance reporter gene are expressed as a fusion with the first three exons of *Lsh* and this hybrid transcript is prematurely terminated at the polyadenylation sequence of the cassette. After induction of ER-Cre with tamoxifen, the Cre recombinase should invert the SA cassette onto an antisense orientation following recombination between loxP and lox511 sites. The SA is not active in the antisense orientation, therefore this inversion leads to normal splicing between the endogenous splice sites (exon 3 to exon 4) allowing

normal expression of the endogenous *Lsh* gene (*Lsh*$^{on}$). FLPo-mediated inversion can further be used to invert the cassette back onto the *Lsh*$^{off}$ orientation (Figure 4.1 A). In this way, a mouse model carrying conditionally reversible *Lsh* alleles (*Lsh*$^{off}$) and tamoxifen-inducible ER-Cre recombinase was generated. Chao Li and Irina Stancheva confirmed the correct integration of the cassette and that the LSH protein is undetectable in *Lsh*$^{off/off}$ ES cells and mice.

Chao Li also designed and optimised a set of primers to detect by PCR the *Lsh*$^{off}$ and *Lsh*$^{on}$ orientation of the GFPneo cassette (Figure 4.1 B). As indicated in Figure 4.1 B, the set of primers consist of three primers P1-P2-P3. The annealing of the primers to different regions depends on the final orientation of the cassette as shown in the schematic diagram (Figure 4.1 B). After PCR amplification the position of the cassette can be distinguished since the PCR product size is different in each case. The amplification product of the cassette in the *Lsh*$^{on}$ position is shorter than the *Lsh*$^{off}$ allowing to distinguish them when the PCR products are run on an agarose gel.

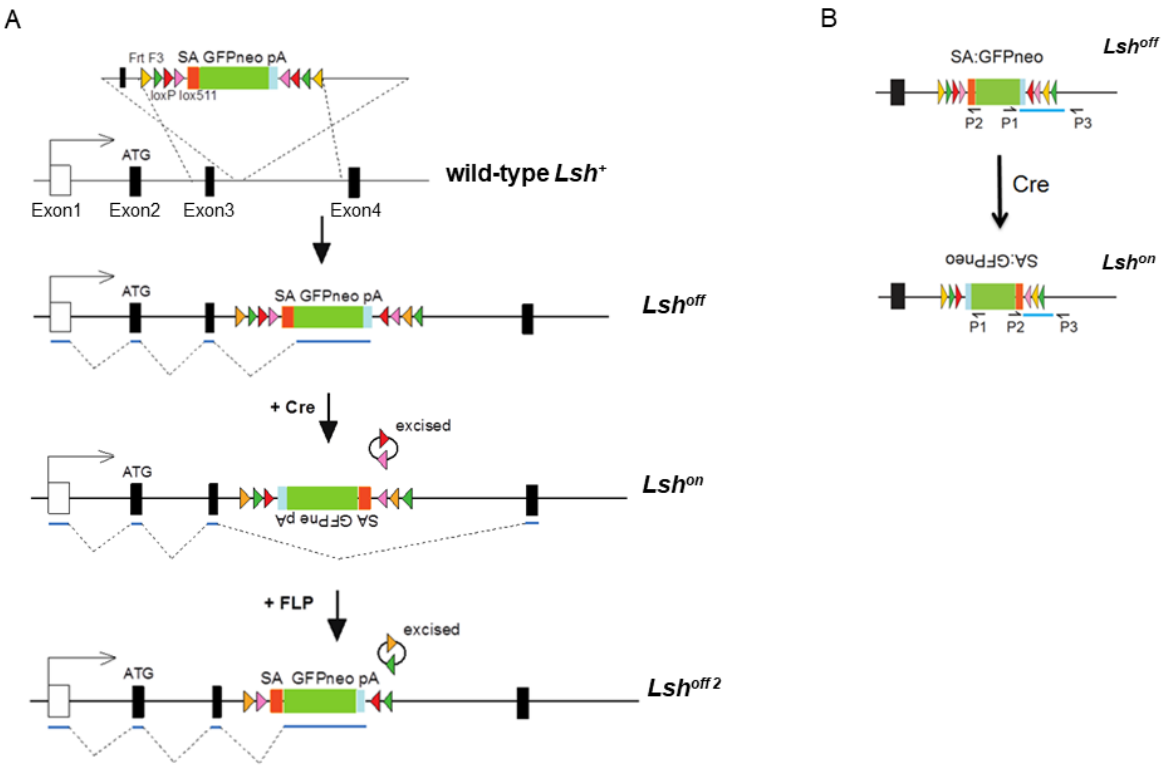


**Figure 4.1. Schematic representation of the conditional reversible *Lsh* allele (*Lsh*$^{off}$) and the mechanism of activation.**

A) GFPneo cassette containing frt (yellow triangles) and F3 (green triangles), heterotypic target sequences for the FLP recombinase; loxP (red triangles) and lox511 (purple triangles), heterotypic target sequences for the Cre recombinase; SA, splice acceptor; GFP,green fluorescent protein; neo, neomycin resistance gene; and pA, polyadenylation sequence; introduced by homologous recombination in the intron 3 of the *Lsh* gene. After translocation of ER-Cre with OHT, the full length *Lsh* gene expression is restored (*Lsh*$^{on}$). FLPo-mediated inversion can further be used to invert the cassette back onto the *Lsh*$^{off}$ orientation. B) Schematic representation of the conversion from *Lsh*$^{off}$ *to Lsh*$^{on}$. P1, P2 and P3 show the annealing of the three primers used for PCR amplification in order to distinguish between the two possible orientations (*Lsh*$^{off}$ or *Lsh*$^{on}$) of the GFPneo cassette. The amplification product from the cassette in *Lsh*$^{on}$ orientation is shorter than the *Lsh*$^{off}$ product allowing to distinguish them when running the PCR products in an agarose gel. Diagram from Irina Stancheva.

Following the establishment of a mouse strain carrying the conditionally reversible *Lsh*$^{off}$ allele and ER-Cre recombinase, I derived mouse embryonic fibroblasts (MEFs) from embryos produced by crosses of heterozygous mice for the *Lsh*$^{off}$ or *Lsh*$^{on}$ allele and ER-Cre recombinase. The MEFs were genotyped using two pairs of primers that amplify the wild-type (*Lsh*$^{+}$) allele and the GFPneo cassette, respectively. The combination of the four primers in a PCR reaction allows to detect in the same PCR reaction run on an agarose gel whether the cells contain the cassette and whether they are homozygous or heterozygous. Also, primers

to detect ER-Cre were used to select cell lines positive for the presence of the recombinase. Finally, another set of primers against the Y chromosome allowed the female or male sex of the MEFs to be determined. Figure 4.2 shows an example of genotyping PCRs for sixteen MEF cell lines immortalised from 13.5 days embryos.

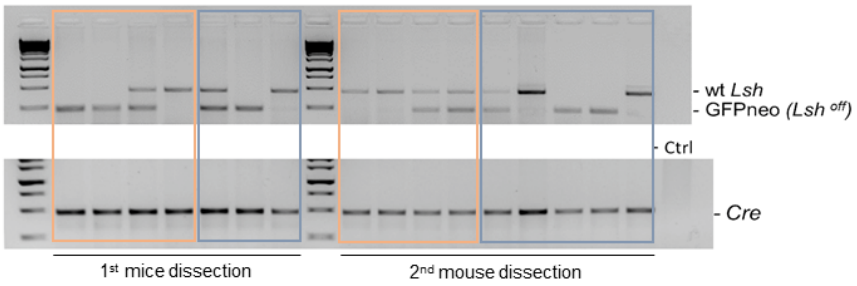


**Figure 4.2. Genotype of sixteen immortalised MEF cell lines derived from 13.5 days embryos.**

gDNA from sixteen different cell lines from two independent dissections was genotyped by PCR for the presence of wt *Lsh* and the GFPneo cassette ($Lsh^{off}$), as well as the presence of ER-Cre recombinase. Ctrl indicates a cell line negative for the presence of ER-Cre recombinase, a negative control for the PCR. The pink rectangles show female genotypes and blue rectangles show male genotypes as determined from PCR for the male-specific SRY gene.

Male cell lines were used in future experiments since some of the analysed loci, such as the *Rhox* cluster of homeobox genes, are on the X chromosome. It is therefore more convenient to use male cell lines because the kinetics of silencing and *de novo* DNA methylation of those genes may differ between the active and the inactive X in female cells. Once the MEF cell lines were derived, genotyped and immortalised with the 3T3 protocol (Sharpless, 2006), characterization and optimisation of the *in vitro* system were required before undertaking further analyses. The characterization of the $Lsh^{off/off}$; ER-$Cre^{+}$ MEFs and the optimisation of cassette inversion in the conditionally reversible *Lsh* knockout cell system will be the focus of this chapter.

## 4.2 Expression of LSH-dependent loci in *Lsh*$^{off/off}$ MEFs

Previously it has been demonstrated in the lab that specific protein coding genes and repetitive elements, such as IAP retrotransposons, gain expression in fibroblasts derived from chromatin remodelling ATPase LSH-null mouse embryos (Myant et al., 2011). Previous studies have also shown expression of Long Terminal Repeats (LTR) retrotransposons and major satellite repeats in the absence of LSH in both MEFs and embryos (Huang et al., 2004).

In order to validate this cellular system, I first had to test whether LSH-dependent loci known to gain expression in the absence of LSH were also misregulated in the *Lsh*$^{off/off}$ MEFs compared to their wild-type *Lsh*$^{+/+}$ counterparts. To do so, I performed qRT-PCRs using cDNA obtained from three independent RNA extractions from *Lsh*$^{off/off}$ and *Lsh*$^{+/+}$ MEFs. Since most of the investigated genes are not expressed in wild-type MEFs, the small variations in noise levels produce a large change in the Ct values obtained by qRT-PCR, the average of wild-type Ct was used as a control reference for the analyses. I determined the expression of repetitive elements (Figure 4.3 A / B) and LSH-dependent protein coding genes including *Gm9*, genes from the *Rhox* cluster, homeobox genes from different clusters, imprinted genes and development and pluripotency associated genes (Figure 4.3 C). The IAP retrotransposons were upregulated 230-fold in the *Lsh*$^{off/off}$ MEFs compared to wild-type cells (Figure 4.3 A). Other repetitive elements such as SINE B1, SINE B2 and MuLV also showed de-repression in *Lsh*$^{off/off}$ MEFs (Figure 4.3 B). However, their expression in the LSH-deficient cells was not as high as the expression of IAPs. The genes of the *Rhox* cluster, *Gm9* gene adjacent to the *Rhox* cluster, *Dppa4* and *Dppa2*, imprinted gene *Peg3* and *Hoxa5* were also highly upregulated in the *Lsh*$^{off/off}$ MEFs, as expected, while *Peg12, Hoxa4* and *Hoxc6* genes showed a small gain of transcription (Figure 4.3 C). The absence of LSH had a higher impact on the expression of pluripotency genes, *Dppa4* and *Dppa2*. Accordingly, previous work had shown that LSH is required for the silencing of development and pluripotency associated genes during differentiation of embryonic stem cells (Myant, Termanis et al. 2011).

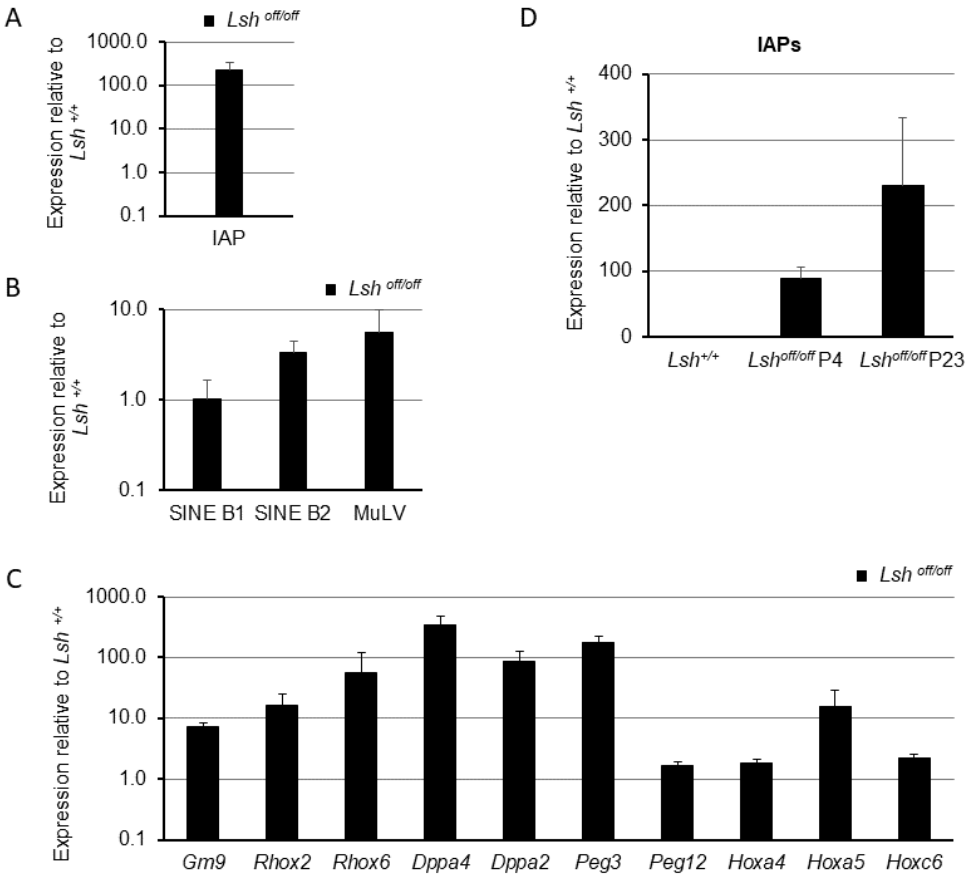


**Figure 4.3. *Lsh*$^{off/off}$ MEFs re-express genes that are normally silenced in the *Lsh*$^{+/+}$ MEFs.**

qRT-PCR analyses show re-expression of repetitive elements and single copy genes in *Lsh*$^{off/off}$ MEFs. The graphs represent average expression +/- SD, normalized to GAPDH, from three independent experiments. All expression values are displayed relative to *Lsh*$^{+/+}$ MEFs. A) Expression of IAP retrotransposons in immortalized *Lsh*$^{off/off}$ MEFs. B) Expression of other repetitive elements in immortalized *Lsh*$^{off/off}$ MEFs. C) Expression of LSH-dependent genes in immortalized *Lsh*$^{off/off}$ MEFs. D) Expression of IAP retrotransposons in primary and immortalized *Lsh*$^{off/off}$ MEFs compared to *Lsh*$^{+/+}$ MEFs.

I also wanted to determine whether the gain of expression in immortalised *Lsh*$^{off/off}$ MEFs (data shown in previous figure Figure 4.3 A / B / C) was similar to the gain of expression in primary cells since there might be a higher misregulation after passaging the cells for a longer period in the absence of LSH. I carried out qRT-PCRs to analyse the expression of IAP retrotransposons in *Lsh*$^{off/off}$ MEFs at passage 4 comparing to the expression in immortalised *Lsh*$^{off/off}$ MEFs at passage number 23. As shown in Figure 4.3 D, the expression for IAPs is higher in immortalised MEFs than primary *Lsh*$^{off/off}$ MEFs, although there is a big variation.

In conclusion, the experiments described above verify the gain of transcription of repetitive elements and previously identified LSH-dependent protein coding genes in *Lsh*$^{off/off}$ MEFs compared to *Lsh*$^{+/+}$ MEFs. This misregulation of LSH-dependent loci in the *Lsh*$^{off/off}$ MEFs

mirrors the changes in gene expression observed in previous studies using *Lsh* knockout cells demonstrating that both approaches used to generate LSH-deficient cells and animals result in very similar alterations in transcriptional regulation.

### 4.3 DNA methylation at LSH-dependent loci in *Lsh*$^{off/off}$ MEFs

Previous work has demonstrated that in the absence of LSH there is loss of DNA methylation from the promoters of genes that are inappropriately expressed in *Lsh* knockout cells as well as other loci that are not upregulated in the knockout cells. LSH is required for normal patterns of DNA methylation at non-repeat sequences genome-wide (Myant et al., 2011; Tao et al., 2011). In addition, other studies have shown that the loci affected by the absence of LSH include repetitive elements and single copy genes (Dennis et al., 2001). This role of LSH in DNA methylation has been further characterized, showing that LSH promotes the establishment of *de novo* DNA methylation (Termanis et al., 2016; Zhu et al., 2006).

Therefore additional validation of the conditional *Lsh* knockout cellular system was to test whether the genes that were aberrantly expressed in *Lsh*$^{off/off}$ MEFs, also lost DNA methylation from their promoter region as shown for the previous *Lsh* knockouts. Bisulfite sequencing analyses of IAP retrotransposons and specific protein coding gene promoters from genomic DNA of both *Lsh*$^{off/off}$ MEFs and *Lsh*$^{+/+}$ MEFs showed that both types of sequence lost DNA methylation in the *Lsh*$^{off/off}$ MEFs. The LTR-embedded promoter of IAP retrotransposons overall lost about 40% of DNA methylation at CpG dinucleotides in *Lsh*$^{off/off}$ MEFs when compared to *Lsh*$^{+/+}$ MEFs (Figure 4.4 A). The loss of DNA methylation at IAPs was not the same for all the sequenced clones. There was an almost complete loss of DNA methylation at half of the clones and no loss of methylation at the rest of clones, which might be explained by some of the IAP retrotransposons acquiring gain of expression while others remained silenced in *Lsh*$^{off/off}$ MEFs. The promoters of *Rhox* genes also showed a significant loss of DNA methylation. The loss of methylation at *Rhox6* and *Rhox9* in *Lsh*$^{off/off}$ MEFs was 55% and 71%, respectively, when compared to *Lsh*$^{+/+}$ MEFs (Figure 4.4 B and C). This data is in agreement with previous work in the lab showing the loss of DNA methylation by 50% at specific LSH-dependent loci including IAPs and *Rhox* genes in cells lacking LSH. A big difference in the methylation levels between *Lsh*$^{off/off}$ and *Lsh*$^{+/+}$ MEFs was also observed at development and pluripotency associated genes, *Dppa2* and *Dppa4*. *Dppa2* promoter lost 84% of methylated CpGs (Figure 4.4 D) and *Dppa4* 68% (Figure 4.4 E) in the *Lsh*$^{off/off}$ MEFs while both promoters were highly methylated in the *Lsh*$^{+/+}$ MEFs. It is important to mention that the calculated loss of DNA methylation at promoters of single copy genes is an average percentage calculated for the analysed clones, but in contrast to IAPs, there is obvious loss in all individual clones.

The higher loss of DNA methylation at the promoters of pluripotency associated genes was accompanied by high gain of expression as shown earlier (Figure 4.4 C). However, this observation is not applicable in all cases since some genes, for example *Rhox6,* that are also highly expressed in *Lsh*$^{off/off}$ MEFs do not lose DNA methylation at such high levels. In order to find a correlation between higher loss of DNA methylation at promoters and higher gene expression I analysed previous expression data obtained from microarrays for *Lsh*$^{-/-}$ MEFs and DNA methylation data obtained from MeDIP sequencing. Again, a correlation could not be found since the expression of each gene is regulated by different transcription factors.

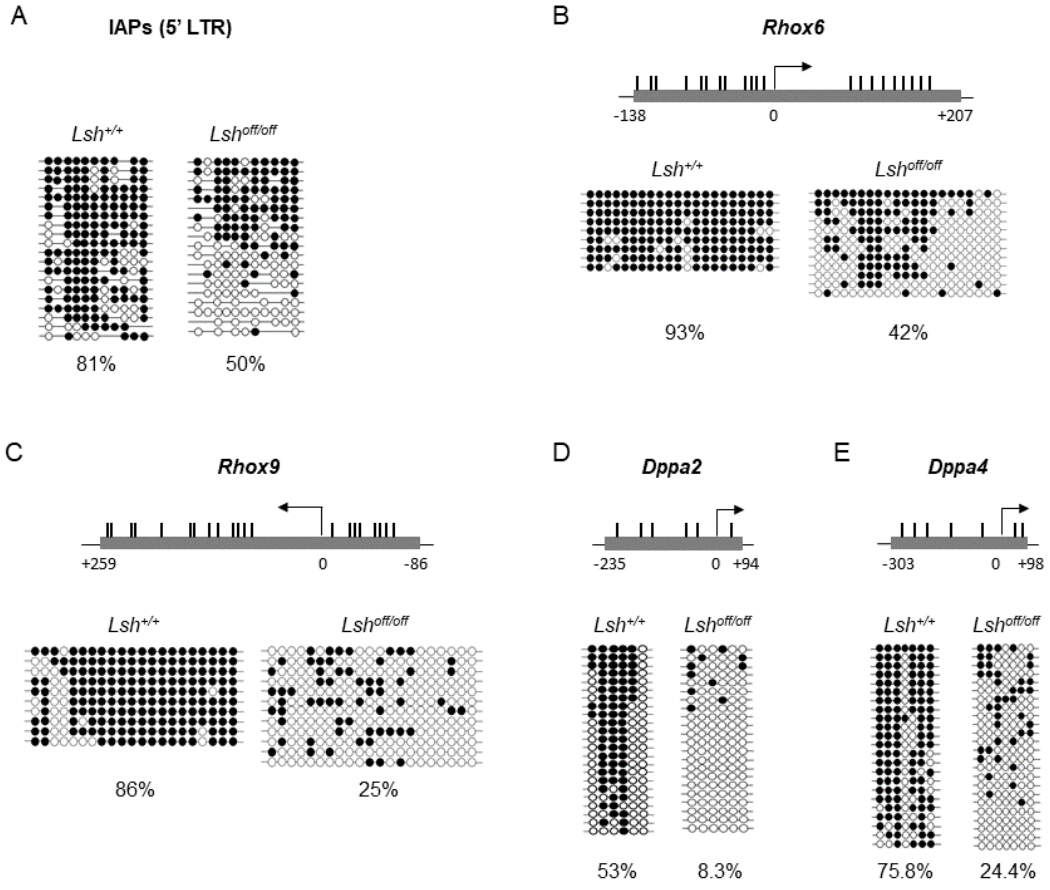


**Figure 4.4. *Lsh*$^{off/off}$ MEFs show hypomethylation at promoters of IAPs and single copy genes.**

Sodium bisulfite DNA sequencing for A) IAPs, B) *Rhox6*, C) *Rhox9*, D) *Dppa2* and E) *Dppa4,* in MEFs at passage 24. The numbers below the promoters graphs indicate bp upstream (-) or downstream (+) from the transcription start site. The black bars on the graphs indicate the distribution of CpG sites analysed. Each row in the diagrams represents a single DNA strand and methylated CpGs are shown as black-filled circles.

Following this, I also tested whether this loss of DNA methylation observed at *Lsh*$^{off/off}$ MEFs varied between primary and immortalised cells comparing the same MEF cell line. As mentioned earlier, LSH is known to play a role in *de novo* DNA methylation, but its role in the maintenance of DNA methylation is not clear yet. IAPs are retrovirus-like repetitive DNA elements that maintain high levels of DNA methylation during development (Rowe & Trono, 2011). To determine if LSH also plays a role in the maintenance of DNA methylation, I decided to analyse whether there was a higher loss of methylation in *Lsh*$^{off/off}$ MEFs at higher passage numbers. Also, we observed that the expression of IAPs was higher in immortalized MEFs than in primary *Lsh*$^{off/off}$ MEFs. This gain of expression could be due to a higher loss of DNA methylation at their promoter region potentially due to impaired maintenance of DNA methylation in the absence of LSH. To address further this hypothesis, I analysed the CpG methylation levels in primary cells at passage 4 and not yet fully immortalised cells at passage 17-18. To better understand any possible variation in DNA methylation at different passages, I also analysed *Lsh*$^{+/+}$ cells as an extra control. The bisulfite sequencing data shown previously in Figure 4.4 was obtained from immortalised MEFs at passage 24-25. Decrease in DNA methylation at the LTR of IAPs could be observed in cells of both genotypes between passage 4 and passage 17. However, the methylation level of IAPs in *Lsh*$^{off/off}$ and *Lsh*$^{+/+}$ MEFs was maintained between cells at passage 17 and immortalised MEFs (Figure 4.5). This could indicate that the variation in methylation is partly due to secondary effects of the immortalization process rather than the higher passage number since the percentages of DNA methylation at passages 17-18 are very similar to the methylation at passage 24-25.

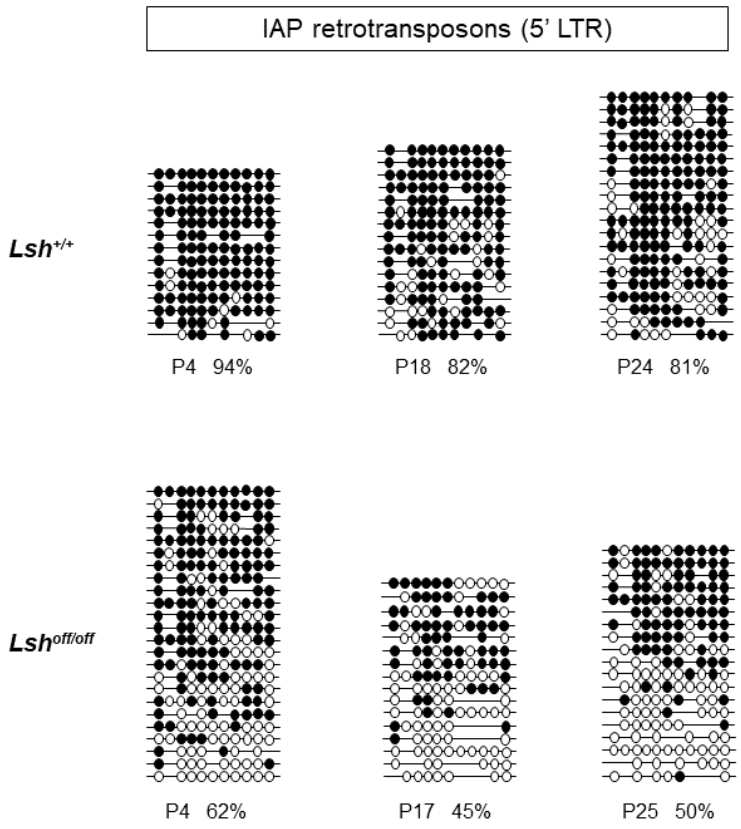


**Figure 4.5. Hypomethylation of IAPs in *Lsh*$^{off/off}$ MEFs is maintained in primary and immortalised MEFs.**

Sodium bisulfite DNA sequencing for IAPs in MEFs at passage 4, passage 17-18 and passage 24-25. DNA methylation in *Lsh*$^{+/+}$ MEFs is shown in the top diagrams and DNA methylation in *Lsh*$^{off/off}$ MEFs is shown in the bottom diagrams. Each row in the diagrams represents a single DNA strand and methylated CpGs are shown as black circles.

In summary, we observed a loss of DNA methylation from promoters of repetitive IAP elements and specific protein coding genes in *Lsh*$^{off/off}$ MEFs. This alteration in DNA methylation at promoters of LSH-dependent loci in *Lsh*$^{off/off}$ MEFs was comparable to the patterns observed in *Lsh*$^{-/-}$ MEFs and thus verifies the role of LSH in the establishment of *de novo* DNA methylation as shown in previous studies. This implies that the *Lsh*$^{off/off}$ MEFs from the conditional *Lsh* knockout can be used for further *in vitro* studies.

## 4.4 Nucleosome occupancy at LSH-dependent loci in *Lsh*$^{off/off}$ MEFs

Recent studies have shown that the role of LSH in the *de novo* DNA methylation at repeat and unique sequences requires intact ATP binding site of the chromatin remodeler (Ren et al., 2015; Termanis et al., 2016). Ren et al also suggested that ATP binding is required for the re-establishment of wild-type patterns of nucleosome occupancy at repeat sequences, which are lost in the *Lsh* knockout cells. The analyses of DNA methylation and nucleosome occupancy was based on an assay known as Nucleosome Occupancy and DNA Methylation (NOMe) sequencing. A NOMe assay carried out to analyse IAPs and Line1 sequences showed that in the absence of LSH both nucleosome occupancy and DNA methylation were lower at the promoters of these loci. When LSH was re-expressed in differentiating embryonic stem cells, it could rescue nucleosome occupancy and DNA methylation at the promoters of repetitive elements. It was hypothesized that ATP-dependent nucleosome remodelling is the primary molecular function of LSH, and that this nucleosome remodelling could promote *de novo* DNA methylation.

To obtain preliminary information of a possible difference in nucleosome density between wild-type MEFs and *Lsh* knockout MEFs, I carried out micrococcal nuclease (MNase) digestion of chromatin from wild-type and knockout cells. Samples were collected after 5, 10, 15, and 20 minutes incubation times; and the digestion products were purified and run in an agarose gel. There were no clear differences in digestion between wild-type and *Lsh*$^{-/-}$ MSCV MEFs. Both cell lines showed a similar pattern of bands after enzymatic digestion for the different incubation times (Figure 4.6 B). The chromatin from *Lsh*$^{-/-}$ MSCV cells seemed to be less accessible to the nuclease as evident from higher molecular weight products at short incubation times. This result contradicts the observation of Ren et al since chromatin seems to be less accessible in our cells lacking LSH while it was shown lower nucleosome occupancy at promoters of cells in the absence of LSH (Ren et al., 2015), which suggests having more accessible chromatin. However, the differences in digestion were not obvious since a small variation in chromatin concentration or incubation time could cause this effect. An hybridization approach, such as Southern Blot, could be used to determine whether the specific loci analysed by Ren et al are less accessible or not to enzymatic digestion in the absence of LSH in our knockout cells. It was determined that 15 minutes of incubation was sufficient to digest the chromatin to mononucleosomal size, around 150 bp digestion product (Figure 4.6 B). The chromatin from wild-type and *Lsh*$^{-/-}$ MSCV MEFs was fully digested in 15 minutes producing a unique band of mononucleosomal size in both cell lines (Figure 4.6 C). The same result was observed for *Lsh*$^{off/off}$ and *Lsh*$^{+/+}$ MEFs, obtaining a similar digestion pattern in both cell lines (Figure 4.6 D). However, small differences in nucleosome density caused by LSH at specific areas of the genome might not be detected by this global approach.

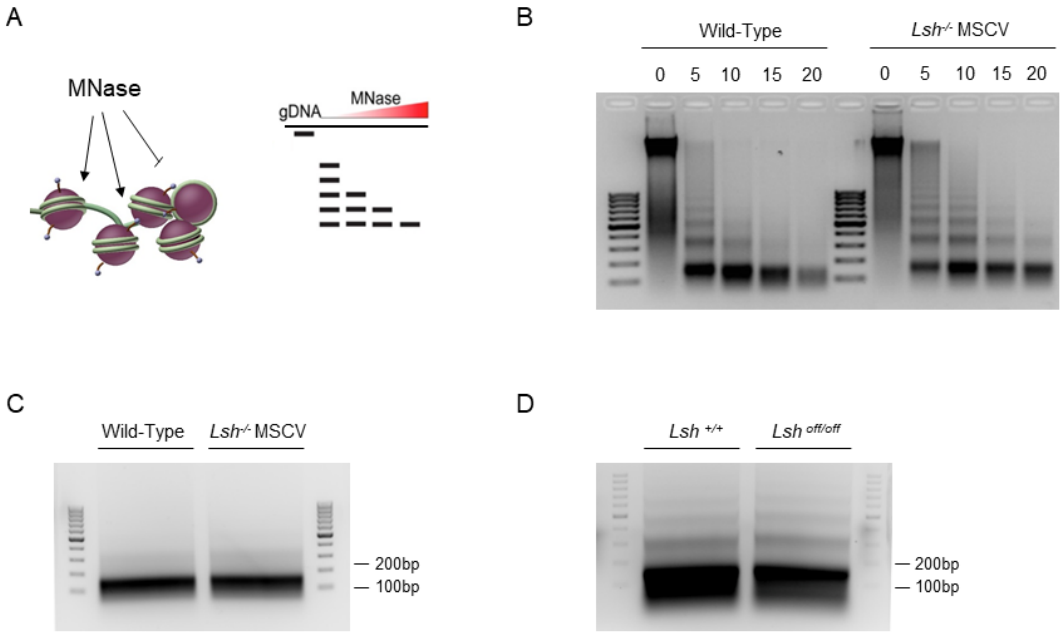


**Figure 4.6. Micrococcal Nuclease (MNase) digestion of chromatin from wild-type and *Lsh* knockout MEFs produces a similar pattern.**

A) Schematic diagram of MNase digestion of chromatin and example of the pattern of bands obtained after running the products of the digestion in an agarose gel. B) MNase products run in a 2% agarose gel after 5, 10, 15 and 20 minutes digestion of chromatin from wild-type and *Lsh*$^{-/-}$ MSCV MEFs with 200 Units of MNase. A non-digested chromatin control is shown as 0 minutes digestion. C) MNase digestion product run in a 2% agarose gel after 15 minutes digestion of chromatin from wild-type and *Lsh*$^{-/-}$ MSCV MEFs with 200 Units of MNase. D) MNase digestion product run in a 2% agarose gel after 15 minutes digestion of chromatin from *Lsh*$^{+/+}$ and *Lsh*$^{off/off}$ MEFs with 200 Units of MNase.

In order to gain insight into the nucleosome occupancy in the *Lsh*$^{off/off}$ MEFs and compare them to the *Lsh*$^{+/+}$ MEFs as well as study any relationship between DNA methylation and nucleosome occupancy, I carried out the NOMe assay for the promoters of *Rhox6* and *Rhox9* genes. We observed loss of DNA methylation by 41% in *Rhox6* and 51% in *Rhox9* (Figure 4.7). This matches the previous result from the bisulfite sequencing. Surprisingly, we did not observe reduced nucleosome occupancy in *Lsh*$^{off/off}$ MEFs compared to *Lsh*$^{+/+}$ MEFs at either of the two promoters. This difference with the previously mentioned study could be due to the different loci analysed, since their study is focused on repetitive elements. Another explanation could be the different cell lines used, since Ren et al worked with differentiating embryonic stem cells and my study was carried out in MEF cell lines.

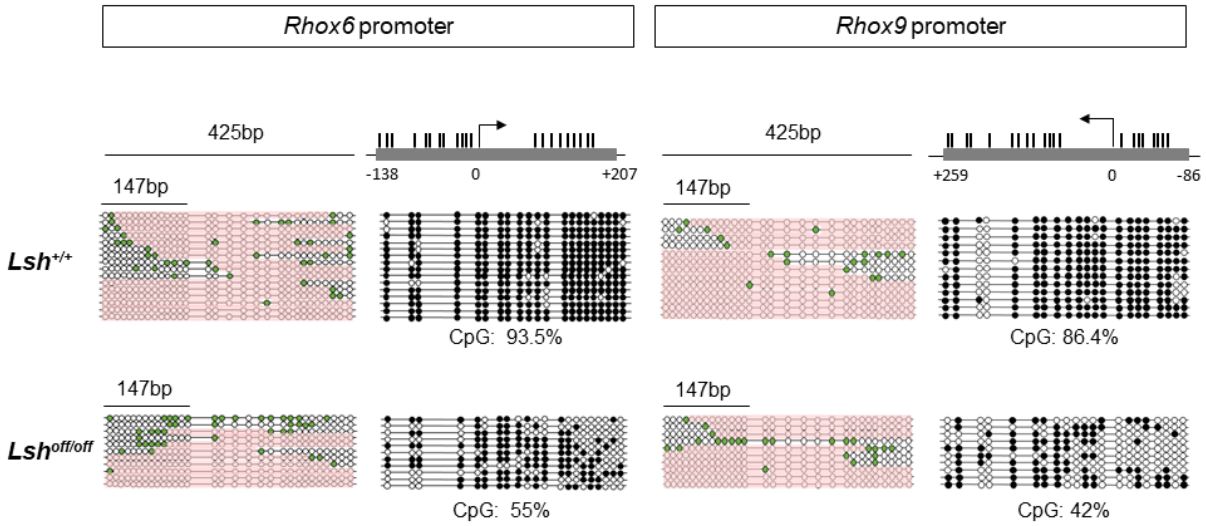


**Figure 4.7. Nucleosome occupancy and DNA methylation at *Rhox6* and *Rhox9* promoters in *Lsh*$^{off/off}$ and *Lsh*$^{+/+}$ MEFs.**

Nucleosome occupancy and DNA methylation (NOMe) assay for *Rhox6* and *Rhox9* promoters in *Lsh*$^{off/off}$ and *Lsh*$^{+/+}$ MEFs. The data for *Lsh*$^{+/+}$ MEFs is shown in the top diagrams and data for *Lsh*$^{off/off}$ MEFs is shown in the bottom diagrams. Each row in a diagram represents a single DNA strand. Nucleosome occupancy diagrams are shown on the left for each promoter, methylated GpCs are shown in green and the area of the promoter where a nucleosome could be positioned is highlighted in pink. Bisulfite sequencing diagrams are shown on the right for each promoter, methylated CpGs are shown in black and the percentage of methylation is indicated below the diagram.

Finally, it is also important to consider the limitation of the NOMe assay when sequencing a small number of colonies and a specific area of a promoter. If a limited number of base pairs is sequenced, a hypothetical nucleosome occupancy can only be inferred since information from the boundary regions is not available. A NOMe-Seq approach might be useful to clarify whether a difference in nucleosome occupancy can also be observed in the *Lsh*$^{off/off}$ MEFs compared to *Lsh*$^{+/+}$ MEFs and whether this difference is specific for repetitive elements or a general feature of all LSH-dependent loci.

## 4.5 Time required for the complete conversion of *Lsh*$^{off}$ allele to *Lsh*$^{on}$

Once the *Lsh*$^{off/off}$ MEFs were characterized, the next step in the establishment of the conditionally reversible *Lsh* knockout cellular system was to test the conversion of the *Lsh*$^{off}$ allele to *Lsh*$^{on}$ upon activation of recombination by ER-Cre. The cells used in this project were genotyped as *Lsh*$^{off/off}$ *ER-Cre*$^{+}$ MEFs and *Lsh*$^{+/+}$ *ER-Cre*$^{+}$ MEFs. *Lsh*$^{+/+}$ *ER-Cre*$^{+}$ MEFs were used as a control for effects that might be caused by the presence of ER-Cre recombinase. The treatment of the cells with 4-hydroxytamoxifen (OHT) should translocate the ER-Cre

recombinase to the nucleus allowing the excision of loxP and lox511 recombinase target sites and inverting the orientation of the cassette from the *Lsh*$^{off}$ to the *Lsh*$^{on}$ position.

Initially, I optimised the concentration of OHT used for treating the cells and the number of days of treatment required to maximise the conversion of the cassette to the *Lsh*$^{on}$ position. Different OHT concentrations were tested, and 600-800nM of OHT seemed the appropriate range of concentration to achieve an optimal conversion of the cassette. The cells were treated with OHT during different time points for a maximum of eight days and all samples were collected on the same day to minimise variation (Figure 4.8 A). After the cells were collected, genomic DNA (gDNA) was extracted and PCR carried out using the primers distinguishing between the *Lsh*$^{off}$ and the *Lsh*$^{on}$ allele (see Figure 4.1 B). In addition, the *Lsh*$^{off}$ allele was analysed by quantitative PCR (q-PCR). I also designed primers to analyse the *Lsh*$^{on}$ allele by q-PCR but none of the pairs of primers could be optimised. These experiments showed that eight days of OHT treatment was sufficient to convert the cassette in most of the cells to the *Lsh*$^{on}$ position (Figure 4.8 B). ImageJ analysis of the PCR (Figure 4.8 C) and the q-PCR detecting the *Lsh*$^{off}$ allele (Figure 4.8 D), showed that eight days of OHT treatment achieved around 90% conversion of the cassette. The treatment was also maintained for a longer period of time to test whether a higher or complete conversion would be obtained, but the results did not indicate further conversion when longer treatments were used (data not shown). Therefore we concluded that eight days of OHT treatment is the time required to obtain an optimal conversion of the GFPneo cassette from *Lsh*$^{off}$ to *Lsh*$^{on}$ orientation.

Once the time required for the conversion of the *Lsh*$^{off}$ allele to *Lsh*$^{on}$ was determined, I analysed whether the protein levels correlated with the 90% efficiency of conversion after eight days of OHT treatment. Western blots showed that the LSH protein levels started to be visible after four days of OHT treatment and became similar to the wild-type levels after eight days of treatment (Figure 4.8 E). However, the LSH expression levels were not completely restored after eight days, so I repeated the OHT treatment and Western blot with three additional biological replicates to assure that during the eight days of treatment the LSH expression and protein levels could be restored to wild-type. We observed that LSH expression levels were restored after the treatment in the three independent replicates (Figure 4.8 F).

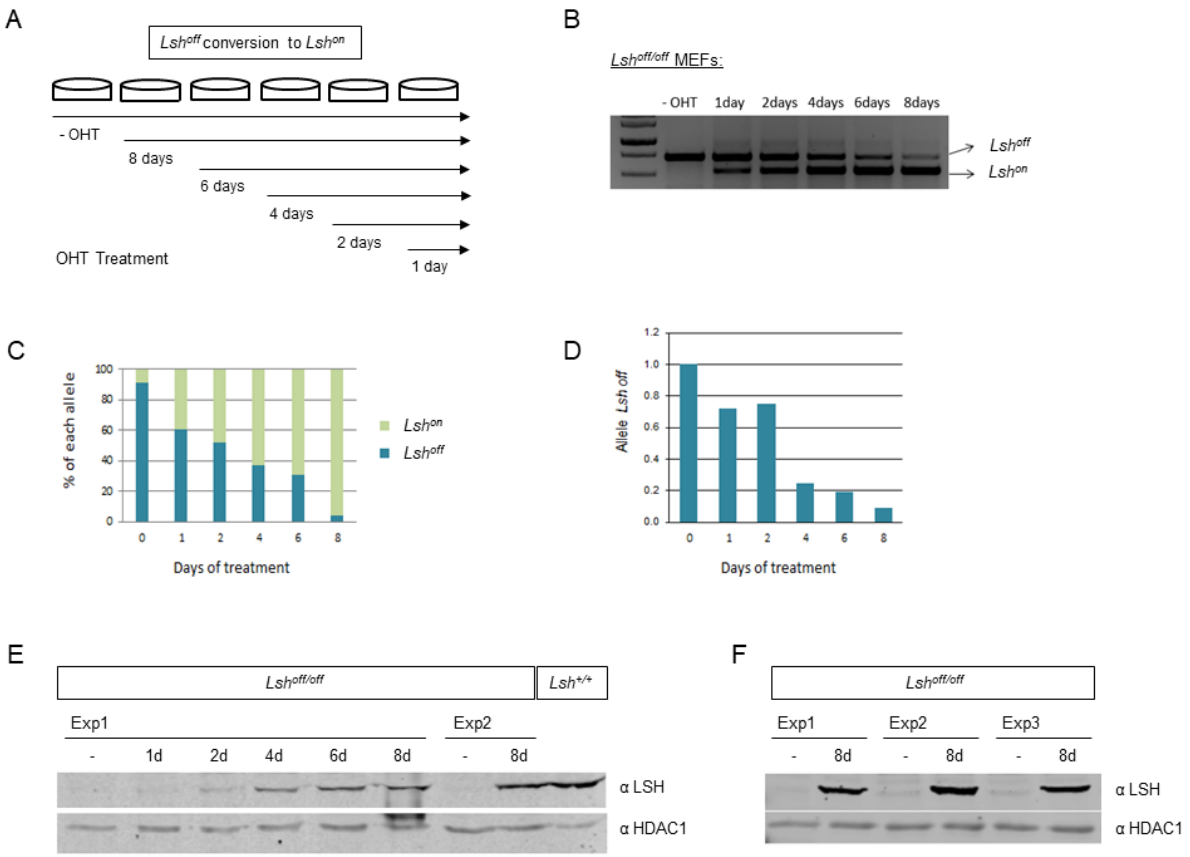


**Figure 4.8. *Lsh*$^{off}$ allele can be converted to *Lsh*$^{on}$ after eight days of OHT treatment restoring LSH protein levels to wild-type.**

A) Schematic diagram showing the OHT treatment to convert the *Lsh*$^{off}$ allele to *Lsh*$^{on}$. B) PCR analyses performed on genomic DNA to monitor the conversion of *Lsh*$^{off}$ allele to *Lsh*$^{on}$ during the 8 days of OHT treatment at final concentration of OHT 600 nM. The conversion efficiency was analysed using a homozygous for *Lsh*$^{off}$ allele cell line. The lower 500 bp band in the gel is the *Lsh*$^{on}$ amplification product while the higher 750 bp band is the *Lsh*$^{off}$ amplification product. C) Quantification of the band intensity performed by ImageJ software. The percentage of each allele is plotted on the graph according to the number of days under OHT treatment. D) q-PCR analyses of *Lsh*$^{off}$ allele using gDNA from each treatment day. The sample without OHT treatment, 0 days, was used as the control for PCR quantification. E) Western blot detecting LSH protein levels from two independent OHT experiments using the same *Lsh*$^{off}$ homozygous cell line. Protein levels are shown at all time points in the first experiment, but only negative control sample, no OHT treatment, and 8 days of OHT treatment sample are shown in the second experiment. Protein level from a control *Lsh*$^{+/+}$ MEF cell line is also shown and HDAC1 nuclear protein is used as a loading control. F) Western blot showing LSH protein levels from three independent OHT experiments using the same *Lsh*$^{off}$ homozygous cell line. Protein levels are shown for the negative control samples, no OHT treatment, and 8 days of OHT treatment samples.

As mentioned earlier in the introduction, the *Lsh*$^{off/off}$ MEFs are neomycin-resistant while the *Lsh*$^{on/on}$ MEFs lose this resistance. When adding geneticin (resistance to geneticin is conferred by the neomycin resistance gene) to the non-treated cells and OHT-treated cells, it was apparent that the colony formation by the non-treated cells was higher than by the OHT-treated MEFs, as expected (Figure 4.9 A and B). This confirms the conversion of the cassette previously tested by PCR and Western blotting. Nevertheless, the colony formation in both

control plates without geneticin should be the same, but less colonies were formed on the plate containing *Lsh*$^{on/on}$ converted MEFs as clearly shown by the quantification (Figure 4.8 B).

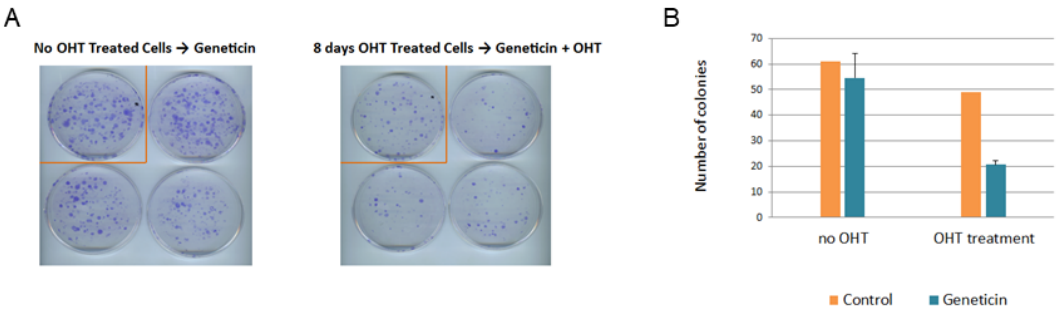


**Figure 4.9. *Lsh*$^{off/off}$ MEFs lose their resistance to geneticin after OHT treatment and conversion to *Lsh*$^{on/on}$.**

A) Colony formation assay using *Lsh*$^{off/off}$ MEFs non-treated and treated with OHT for a period of 8 days. Cells were plated in a low density and the medium was supplemented with geneticin to final concentration of 250 µg/ml. Medium added to plates containing 8 days OHT treated cells was also supplemented with OHT during the three weeks of geneticin treatment. The orange square indicates the negative control plates without geneticin treatment. B) The graph represents the average number of colonies detected on geneticin treated plates in non-treated and OHT-treated cells. The number of colonies in the negative control plates without geneticin is also shown.

To summarize, eight days of OHT treatment is sufficient to convert the *Lsh*$^{off}$ allele to *Lsh*$^{on}$ in immortalised MEFs. This conversion also restored LSH expression in the *Lsh*$^{on/on}$ MEFs to wild-type levels. The conversion of the cassette is also confirmed by the reduced resistance to geneticin after eight days of OHT treatment. All these approaches show that the system functions as expected and indicate that the reversible *Lsh* knockout cells can be used for further studies.

## 4.6 Reproducibility of *Lsh*$^{off}$ to *Lsh*$^{on}$ allele conversion and LSH expression in MEFs

One of my aims was to determine the order of events leading to gene silencing when LSH expression is restored in the *Lsh* knockout MEFs and the time required for acquisition of gene silencing and DNA methylation. In order to study this, time point experiments after OHT treatment are required to understand what changes in chromatin come first and how much time is needed for the establishment of gene silencing. In the previous section I demonstrated that LSH protein levels are restored after eight days of OHT treatment, but the stability of LSH expression over time was not tested. This section will focus on the analysis of LSH expression

over time after OHT treatment and the reproducibility of the conversion of the cassette to the *Lsh*$^{on}$ orientation.

After eight days of OHT treatment, the converted *Lsh*$^{on/on}$ MEFs were maintained in culture for a period of four weeks without renewal of OHT treatment. Surprisingly, LSH protein levels in these long-term experiments decreased after four weeks to levels similar to the non-treated samples (Figure 4.10 A). When genomic DNA extracted from this experiment was used to analyse the cassette orientation during this period we observed that the intensity of the *Lsh*$^{on}$ band in the agarose gel decreases as the number of weeks during which the cells were cultured increases (Figure 4.10 B). This decrease of the *Lsh*$^{on}$ cassette is accompanied by an increase of the *Lsh*$^{off}$ cassette. Since the *Lsh*$^{on}$ cassette cannot be converted back to *Lsh*$^{off}$ orientation, we hypothesized that the small percentage of non-converted cells after treating with OHT (about 10-20% of the cells) could be outgrowing the *Lsh*$^{on/on}$ converted MEFs. This could explain the decrease in the LSH protein levels after four weeks. A second hypothesis could be that the *Lsh*$^{on/on}$ MEFs are suddenly dying while the *Lsh*$^{off/off}$ MEFs survive and outgrow them. However, there was no visual difference in the number of dead cells between the OHT-treated and non-treated plates so this explanation was discarded. To confirm this, I carried out growth rate measurement to determine the population doubling time of each cell population, the non-treated MEFs and the OHT-treated cells. I tried to study the doubling time of the *Lsh*$^{off/off}$ and *Lsh*$^{on/on}$ MEFs, but it was not possible to select for the *Lsh*$^{on/on}$ cells, so I ended up with a mixed cell population and could not determine whether the non-converted cells were outgrowing the *Lsh*$^{on/on}$ MEFs. OHT was not added to the medium during the period of four weeks when the cells were maintained in culture. Continuous and/or discontinuous treatment with OHT during this period could solve the problem of the non-converted cells outgrowing the converted population. Therefore the cells converted to the *Lsh*$^{on}$ orientation after eight days of OHT treatment were cultured for four weeks either with or without OHT renewal. The position of the cassette after four weeks with the different OHT treatments was analysed by PCR. However, maintaining the cells for a longer period under OHT treatment did not change the outcome. Both, continuous and discontinuous treatment with OHT resulted in expansion of *Lsh*$^{off/off}$ MEF population after four weeks of treatment (Figure 4.10 C). By the third week maintaining the cells, the LSH protein was not detectable anymore by Western blot (Figure 4.10 D). Nevertheless, discontinuous treatment with OHT kept a higher expression of the protein after two weeks (Figure 4.10 D). The only possible explanation at this point, is that the initially non-converted MEFs (10-20%), which are either resistant to OHT treatment or may have acquired a mutation that prevents conversion of the cassette, outgrow the converted *Lsh*$^{on/on}$ MEFs even under continuous OHT treatment.

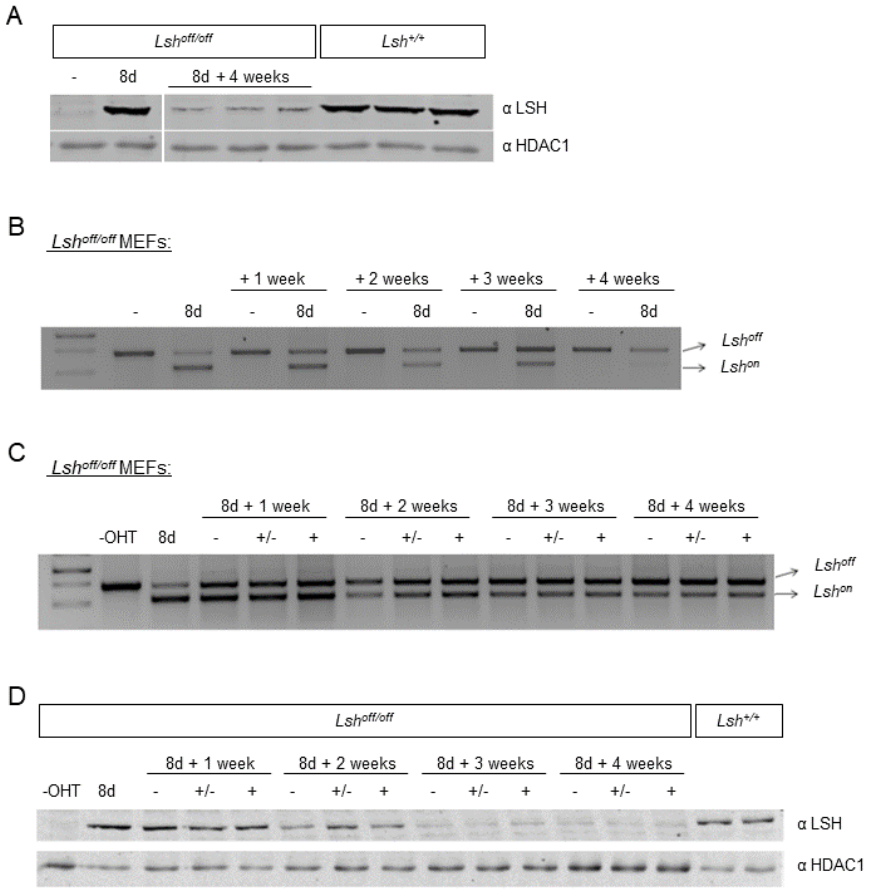


**Figure 4.10. LSH protein levels are restored to wild-type after tamoxifen treatment but there is loss of the protein over a period of four weeks while the *Lsh*$^{off/off}$ cells outgrow *Lsh*$^{on/on}$ MEFs.**

A) Western blot showing LSH protein levels after OHT treatment in *Lsh*$^{off}$ homozygous cell line. Protein levels are shown for the negative control samples, no OHT treatment, and at 8 days of OHT treatment. 8d + 4weeks, represents a triplicate of the 8 day OHT treated cells that were cultured longer, for a period of four extra weeks, without renewal of OHT treatment. *Lsh*$^{+/+}$ shows a triplicate of the protein levels from a wild-type MEF cell line. HDAC1 nuclear protein was used as a loading control. B) PCR analyses performed on genomic DNA monitor the conversion of *Lsh*$^{off}$ allele to *Lsh*$^{on}$ during the 8 days of OHT treatment. The conversion efficiency was analysed using a cell line homozygous for the *Lsh*$^{off}$ allele. The non-treated and 8 day OHT-treated cells were cultured longer, for a period of four extra weeks, without renewal of OHT treatment. Samples were collected each week to analyse the presence of the converted cassette. The lower 500 bp band in the gel is the *Lsh*$^{on}$ amplification product while the higher 750 bp band is the *Lsh*$^{off}$ amplification product. C) PCR analyses performed on genomic DNA monitor the conversion from *Lsh*$^{off}$ allele to *Lsh*$^{on}$ after 8 days of OHT treatment compared to non-treated cells. The 8 day OHT treated cells were cultured longer, for a period of four extra weeks, without renewal of OHT (-), with discontinuous renewal of OHT (+/-) and with continuous OHT treatment (+). Samples were collected each week to analyse the presence of the converted cassette under the different conditions of OHT treatment. D) Western blot showing LSH protein levels after OHT treatment in *Lsh*$^{off}$ homozygous cell line. Protein levels are shown for the negative control samples, no OHT treatment, and at 8 days of OHT treatment. The 8 day OHT treated cells were cultured for four extra weeks without renewal of OHT (-), with discontinuous renewal of OHT (+/-) and with continuous OHT treatment (+). Samples were collected each week to analyse the protein levels after the different conditions. *Lsh*$^{+/+}$ shows a duplicate of the protein levels from a wild-type MEF cell line. HDAC1 nuclear protein was used as a loading control.

Another issue became apparent when the *Lsh*$^{off/off}$ MEFs were treated again with OHT to reproduce the conversion of the cassette. The conversion obtained after eight days of OHT treatment was not as efficient as in earlier experiments. We observed that only about 50% of the cassette was converted to the *Lsh*$^{on}$ band (Figure 4.11 A). The initial conversion rate obtained in the first experiments was about 90%. To try solving the issue, the concentration of OHT was increased to 780 nM and a new batch of OHT was purchased to avoid stability problems. Disappointingly, the initial conversion rate could not be reproduce anymore (Figure 4.11 B). To confirm that there was a problem with the conversion and not with the PCR detecting the cassette orientation, Western blots were performed to detect LSH protein levels. The protein levels obtained after eight days of OHT treatment were lower than the levels from the wild-type cells and the converted *Lsh*$^{on/on}$ MEFs in the initial experiments (Figure 4.11C).

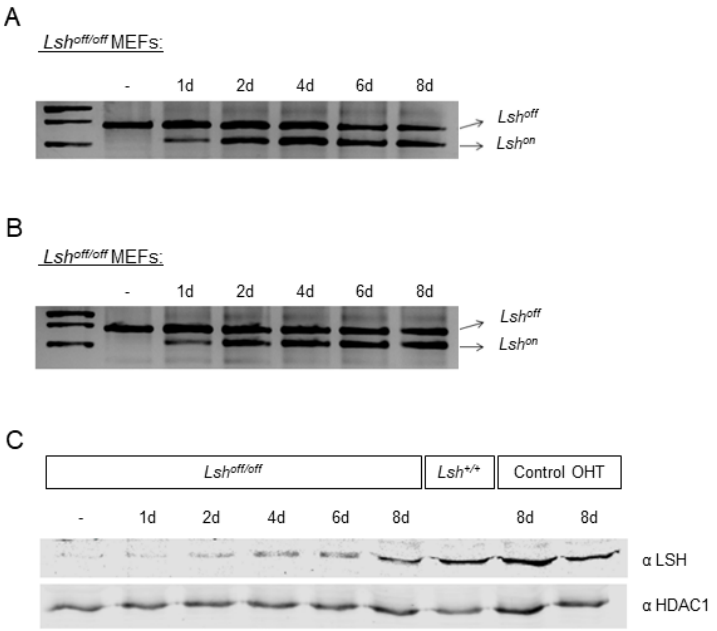


**Figure 4.11. The conversion of *Lsh*$^{off}$ to *Lsh*$^{on}$ allele and LSH expression are not reproducible in immortalised MEFs.**

PCR analyses of gDNA extracted from *Lsh*$^{off/off}$ MEFs at different time points during the eight days of OHT treatment. Negative control sample without OHT treatment (-) is also shown. The cells were treated with OHT at final concentration of 600 nM (A) and 780 nM (B). The 500 bp band is the *Lsh*$^{on}$ amplification product while the higher 750 bp band is the *Lsh*$^{off}$ amplification product. C) Western blot showing LSH protein levels from the treatment experiment using 780 nM concentration of OHT. Protein levels are shown for all time points and 8 days of OHT treatment samples from the previous experiments are used as positive controls. *Lsh*$^{+/+}$ shows the protein level in a wild-type MEF cell line. HDAC1 nuclear protein used as a loading control.

The capacity to select for the *Lsh*$^{on/on}$ MEFs would have been very useful even with a very low rate of conversion, since it would avoid having a mixed population of cells in the experiments and would solve the issue of the unconverted cells outgrowing the *Lsh*$^{on/on}$ MEFs. Therefore I designed a cassette containing a puromycin resistance gene in the antisense orientation flanked by the recombinase target sites loxP and lox511 (heterotypic target sequences for the Cre-recombinase). This design was similar to the design of the GFPneo cassette. The cassette was cloned into a pMSCV hygromycin vector, lentivirus containing the vector were generated and the *Lsh*$^{off/off}$ MEFs were transduced. Ideally, when treating the cells with OHT to convert from *Lsh*$^{off}$ to the *Lsh*$^{on}$ orientation, the puromycin cassette should also be converted to the right orientation, producing resistance to puromycin in the same cells that have converted the GFPneo cassette. The *Lsh*$^{on/on}$ MEFs could now be selected by treating the cells with puromycin (see diagram in Figure 4.12 A). I designed different pairs of primers to detect the conversion of the puromycin cassette by PCR, but none of the pairs could be optimised. In order to test if the puromycin cassette was working, the *Lsh*$^{off/off}$ MEFs were treated again with OHT for a period of eight days and then the cells were treated with puromycin to select for the *Lsh*$^{on/on}$ MEFs. This approach was assuming that both cassettes were converted in the same cells. However, the cells did not survive the puromycin treatment. To test whether the cassettes were converted, I analysed the conversion of the GFPneo cassette of two independent experiments after eight days of OHT treatment (Figure 4.12 B). The efficiency of conversion to the *Lsh*$^{on}$ orientation was very low. Probably, the efficiency of conversion for the puromycin cassette was similar, explaining why the cells were not resistant to puromycin treatment. To increase the conversion of both cassettes, a new vector containing a more efficient inducible Cre-recombinase, Cre-ER$^{T2}$, was generated. Cre-ER$^{T2}$ contains a triple mutation of the ligand-binding domain of the estrogen receptor (ER) to decrease the nuclear translocation in the absence of the inducer tamoxifen, while increasing the sensitivity to tamoxifen and the recombination efficiency (Indra et al., 1999). The vector containing Cre-ER$^{T2}$ was introduced into the *Lsh*$^{off/off}$ MEFs by lentiviral transduction. This newly generated cell line was used to repeat the OHT treatment for eight days. The efficiency of conversion with the new Cre-ER$^{T2}$ was very high, producing similar results to the initial OHT treatments (Figure 4.12 C). This high efficiency of cassette conversion combined with the possibility to select for the converted cells with puromycin could be the solution to the issue of non-converted cells outgrowing the *Lsh*$^{on/on}$ MEFs. Cre-ER$^{T2}$ was cloned into a pMSCV hygromycin vector, lentivirus were generated and the *Lsh*$^{off/off}$ MEFs were transduced with lentivirus containing Cre-ER$^{T2}$ and the puromycin cassette. The newly generated cell line was then treated with OHT for eight days and then puromycin treatment was applied to select for the *Lsh*$^{on/on}$ MEFs (Figure 4.12 D). However, the cells again did not survive the puromycin treatment. There must had been a problem in the design of the puromycin cassette that

impaired the appropriate expression of puromycin and thus could explain why the converted cells did not survive the antibiotic selection treatment.

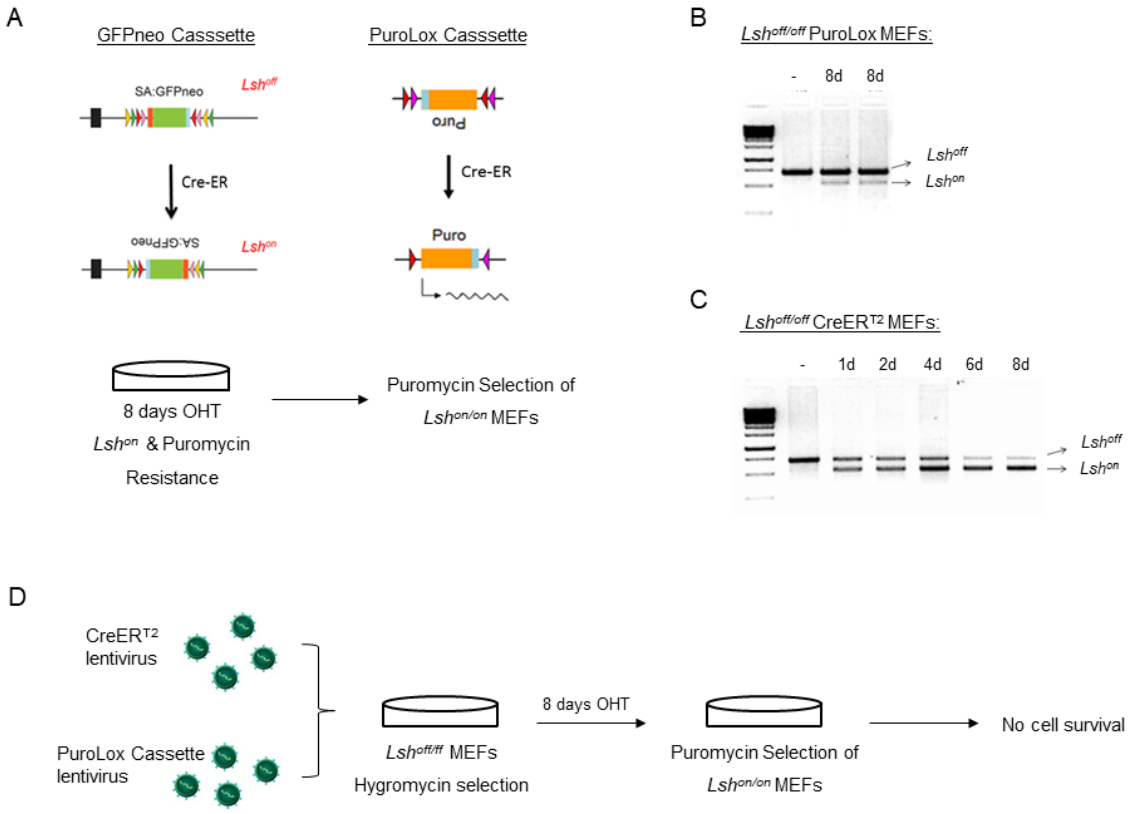


**Figure 4.12. Selection of *Lsh*$^{on/on}$ MEFs by puromycin treatment and the use of Cre-ER$^{T2}$.**

A) Schematic diagram of the GFPneo and PuroLox cassettes to select for *Lsh*$^{on/on}$ MEFs after eight days of OHT treatment. B) PCR analyses of gDNA extracted from *Lsh*$^{off/off}$ PuroLox MEFs after eight days of OHT treatment. Samples are from cells treated for eight days with OHT in two independent experiments. Negative control sample without OHT treatment (-). The 500 bp band is the *Lsh*$^{on}$ PCR amplification product while the 750 bp band is the *Lsh*$^{off}$ amplification product. C) PCR analyses of gDNA extracted from *Lsh*$^{off/off}$ Cre-ER$^{T2}$ MEFs at different time points during the eight days of OHT treatment. Negative control sample without OHT treatment (-) is also shown. The 500 bp band is the *Lsh*$^{on}$ PCR amplification product while the 750 bp band is the *Lsh*$^{off}$ amplification product. D) Schematic diagram showing the approach to select the *Lsh*$^{on/on}$ MEFs with puromycin after tamoxifen treatment of *Lsh*$^{off/off}$ cells transduced with lentivirus containing Cre-ER$^{T2}$ and PuroLox cassette.

To understand why the efficiency of conversion was becoming lower in the immortalised *Lsh*$^{off/off}$ MEFs, the same cell line with increasing passage numbers was treated with OHT for seven days. Surprisingly, I found that cells with higher passage numbers showed decrease in the efficiency of conversion (Figure 4.13 A). There was an almost complete conversion to *Lsh*$^{on}$ at passage 3 after seven days of tamoxifen treatment. This efficiency decreased at passage 11 and was even lower at passage 23. The use of primary cell lines would circumvent the problems with cassette conversion, although it would require more embryo dissections to be able to obtain enough material to carry out further experiments. Primary MEF cell lines were treated with tamoxifen to test the reproducibility of the conversion of the cassette and samples were collected after seven, ten and fourteen days of treatment. The conversion of the cassette was almost complete after seven days, and this conversion was stably maintained over the fourteen days of treatment (Figure 4.13 B). However, LSH protein levels were not detectable by Western blot in any of the collected samples, not even in the *Lsh*$^{+/+}$ pMEFs (Figure 4.13 C). The *Lsh*$^{+/+}$ pMEFs had been maintained in culture without OHT treatment for the same period to be collected together with the treated *Lsh*$^{off/off}$ pMEFs. The cells were not dividing when these samples were collected, since primary MEFs stop proliferating after a few passages in culture when they enter replicative senescence. Previous studies have shown that the expression of LSH is proliferation-associated (Raabe et al., 2001), since LSH is highly expressed in proliferating tissues such as thymus and testis (Jarvis et al., 1996; D. W. Lee et al., 2000). Moreover, E2F1 has been shown to play a crucial role in transcriptional control of the *Lsh* gene. The decrease in E2F1 expression during senescence leads to lower expression of LSH (Niu et al., 2011). This association with proliferation explains why LSH is undetectable in the samples collected from primary MEFs entering replicative senescence, including the *Lsh*$^{+/+}$ MEFs.

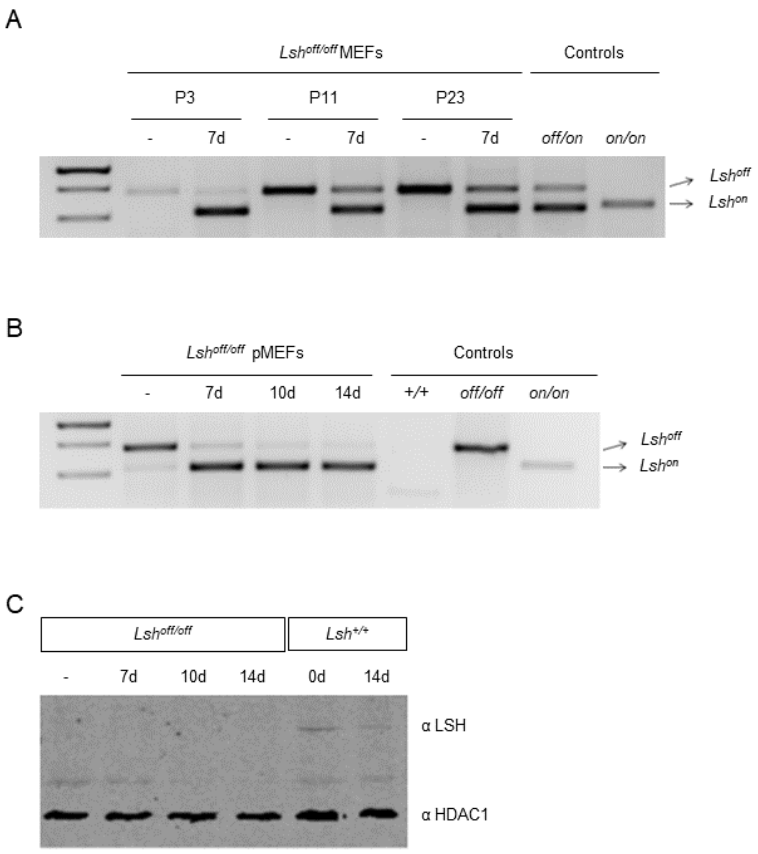


**Figure 4.13. *Lsh*$^{off}$ to *Lsh*$^{on}$ conversion works efficiently in primary MEFs but LSH is not express in non-proliferating cells.**

A) PCR analyses of gDNA extracted from *Lsh*$^{off/off}$ MEFs after seven days of OHT treatment carried out with the same cell line at different passage numbers. Negative control sample without OHT treatment (-). *Lsh*$^{off/on}$ and *Lsh*$^{on/on}$ controls are shown. The 500 bp band is the *Lsh*$^{on}$ PCR amplification product and the 750 bp band is the *Lsh*$^{off}$ amplification product. B) PCR analyses of gDNA extracted from *Lsh*$^{off/off}$ primary MEFs after different days of OHT treatment. Negative control sample without OHT treatment (-). *Lsh*$^{+/+}$, *Lsh*$^{off/off}$ and *Lsh*$^{on/on}$ controls are shown. The 500 bp band is the *Lsh*$^{on}$ PCR amplification product and the 750 bp band is the *Lsh*$^{off}$ amplification product. C) Western blot detecting LSH protein levels from the OHT treatment experiment using primary *Lsh*$^{off/off}$ MEFs. Non-treated samples (-). Protein levels are shown for all time points and *Lsh*$^{+/+}$ samples at the beginning (0d) and end (14d) of the experiment are used as positive controls. HDAC1 nuclear protein was used as a loading control.

Since the conversion of the GFPneo cassette works better in early passage cells, a new approach to immortalise cells quickly after derivation was used. The 3T3 immortalization protocol that I used initially, requires up to four months to immortalise the cell lines. A quicker and less variable method to immortalise primary cells consist on lentiviral transduction of cells with virus containing the Large T antigen (LTag) of SV40 which inactivates p53 transcription factor (Sharpless, 2006). Retrovirus containing pBabe Large T antigen vector were kindly provided by Sara Buonomo and used to immortalise primary *Lsh*$^{off/off}$ and *Lsh*$^{+/+}$ MEFs. The OHT treatment was repeated using these new immortalised cell lines, and gDNA and protein

samples were collected at different time points to analyse the conversion of the cassette and the expression of LSH protein. After seven days of OHT treatment, most of the cassette was converted to *Lsh*$^{on}$ orientation and this conversion was stably maintained over the fourteen days of treatment (Figure 4.14 A). The LSH protein was also detectable by Western blot after seven days of treatment and the expression became higher after ten days of treatment, reaching similar levels of expression to that in *Lsh*$^{+/+}$ MEFs (Figure 4.14 B). However, when the OHT treated cells where maintained in culture over fourteen days, with or without continuous OHT treatment, the non-converted cells started outgrowing the *Lsh*$^{on/on}$ MEFs (data not shown). To maintain the cells with the allele converted to *Lsh*$^{on}$ orientation during the fourteen days period, heat inactivated foetal bovine serum was used since *Lsh*$^{on/on}$ MEFs seemed to proliferate better when the medium was supplemented with this serum. The gain of expression for LSH-dependent loci and loss of DNA methylation at their promoters was analysed for the *Lsh*$^{off/off}$ LTag immortalised MEFs (data not shown). The misregulation of both, appropriate gene silencing and DNA methylation, was similar to the changes observed in spontaneous immortalised *Lsh*$^{off/off}$ MEFs, validating the use of these cell lines for future experiments.

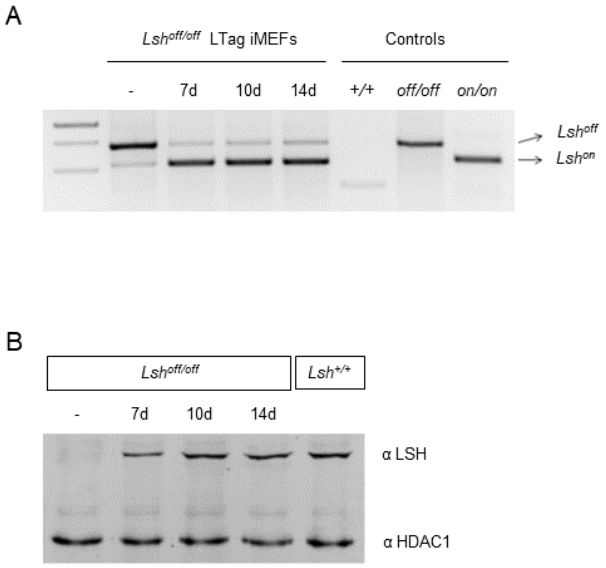


**Figure 4.14. *Lsh*$^{off}$ to *Lsh*$^{on}$ conversion works efficiently in Large T antigen (LT-ag) immortalised MEFs and LSH is expressed at wild-type levels.**

A) PCR analyses of gDNA extracted from *Lsh*$^{off/off}$ LTag immortalised MEFs after seven, ten and fourteen days of OHT treatment. Negative control sample without OHT treatment (-). *Lsh*$^{+/+}$, *Lsh*$^{off/off}$ and *Lsh*$^{on/on}$ controls are shown. The *Lsh*$^{on}$ and *Lsh*$^{off}$ amplification products are indicated. C) Western blot showing LSH protein levels in the OHT treated *Lsh*$^{off/off}$ LTag MEFs. Non-treated samples (-). Protein levels are shown for all time points and *Lsh*$^{+/+}$ sample is used as a positive control. HDAC1 nuclear protein was used as a loading control.

In summary, seven days of OHT treatment is sufficient to convert the *Lsh*$^{off}$ allele to *Lsh*$^{on}$ in SV40 Large T antigen immortalised MEFs. After the conversion of the cassette, the LSH expression is detectable in proliferating *Lsh*$^{on/on}$ MEFs and comparable to endogenous wild-type expression levels. The *Lsh*$^{on/on}$ MEFs population and the expression of LSH protein are stably maintained for up to fourteen days of treatment. This time frame might be sufficient to determine the time required for acquisition of gene silencing and DNA methylation when LSH is reintroduced into the *Lsh* knockout MEFs and follow the order of events leading to gene silencing.

## 4.7 Discussion

The characterization of the *Lsh*$^{off/off}$ MEFs shown in this chapter is in agreement with previous studies demonstrating that LSH deficiency affects the expression and DNA methylation at promoters of specific single-copy protein coding genes (Myant et al., 2011; Tao et al., 2011; Xi et al., 2007) and repetitive elements (Dennis et al., 2001; Huang et al., 2004). However, the gain of expression observed at *Lsh*$^{off/off}$ MEFs is not the same for all LSH-dependent loci in study. *Rhox* genes gain expression and lose DNA methylation from their promoters in the absence of LSH, demonstrating that LSH contributes to their silent state. Pluripotency genes are silenced upon differentiation of stem cells, but they are highly expressed in *Lsh*$^{off/off}$ cells, as well as the imprinted *Peg3* gene. The *Hox* genes are not that highly expressed, showing gain of transcription similar to *Gm9* and another imprinted gene *Peg12*. The promoters of LSH-dependent genes lose DNA methylation by at least 55% in *Lsh*$^{off/off}$ MEFs compared to *Lsh*$^{+/+}$ MEFs. The overall loss of methylation at IAPs is 40%, including some sequences which are completely unmethylated and some sequences that barely show loss of DNA methylation, which might be explained by some of the IAP retrotransposons gaining expression while others remained silenced in *Lsh*$^{off/off}$ MEFs, as mentioned earlier. Importantly, a higher loss of DNA methylation is not always accompanied by gain of gene expression. Other factors besides LSH, such as transcription factors which may not be present in MEFs, could also play an important role in gene expression. Moreover, restoration of LSH expression in *Lsh* knockout cells is not sufficient to control the expression of all misregulated genes since additional exogenous signals are required for silencing of development and pluripotency associated genes (Termanis et al., 2016). The difference in the loss of DNA methylation at IAPs observed between primary cells and immortalised cells was similar in *Lsh*$^{off/off}$ and *Lsh*$^{+/+}$ MEFs. This could indicate that the variation in methylation between cells at passage 4 and passage 17-18 is due to secondary effects of the immortalization process since the cells need to escape senescence. Culture of primary cells

is limited to a certain number of cell divisions before the cells enter replicative senescence, which happens around passage 5 in the primary MEFs used in this study. Many studies have shown that replicative senescence is associated with changes in DNA methylation at specific CpG dinucleotides and the levels of DNA methylation can either increase or decrease over subsequent passages (Bork et al., 2010; Koch et al., 2013), although DNA hypomethylation is more common (Cruickshanks et al., 2013). Also, these changes in methylation appear to be acquired stochastically (Franzen et al., 2017). The percentage of DNA methylation at passages 17-18 is very similar to the methylation at passage 24-25 for both $Lsh^{off/off}$ and $Lsh^{+/+}$ cells, indicating that the loss of methylation observed from passage 4 to passage 17-18 is not related to higher passage numbers, but to the primary cells entering replicative senescence. This suggests that LSH might not play a role in maintaining DNA methylation. Many studies have shown the role of LSH in the establishment of DNA methylation but a role in the maintenance of DNA methylation has not been clearly determined yet (Briones & Muegge, 2012). However, some *in vivo* data (unpublished) generated in the lab suggests that LSH might support maintenance as well as *de novo* DNA methylation. More loci should be analysed including single copy genes and other experiments should be carried out in order to clarify whether LSH has a role in the maintenance of DNA methylation. The best experiment to analyse this would have been to convert the allele in $Lsh^{on/on}$ cells to the $Lsh^{off}$ orientation and study whether the normal patterns of DNA methylation characteristic of wild-type cells are lost in the absence of LSH. This would prove that LSH has a role in the maintenance of DNA methylation. I tried to perform this experiment but different issues arose and the conversion could not be optimised.

The conditionally reversible *Lsh* knockout cellular system seemed to be producing an efficient conversion of the $Lsh^{off}$ allele to $Lsh^{on}$ orientation, as expected, and resulted in appropriate expression of LSH protein with levels similar to those in wild-type cells. However, several issues arose when the converted cells were grown for longer periods of time in culture. The design of the cassette includes GFP and neomycin resistance as a positive selection for the $Lsh^{off/off}$ MEFs, but it is a lacking a selection for the converted $Lsh^{on}$ allele. This lack of selection for the converted cells would not have been a problem if the $Lsh^{off/off}$ cells were not outgrowing the $Lsh^{on/on}$ cells in the mixed population in a short period of time. Also, GFP could have been used to indirectly select for the $Lsh^{on/on}$ cells by cell sorting, but unluckily it is not fluorescent when directly coupled with neomycin. The doubling time of the $Lsh^{off/off}$ and $Lsh^{on/on}$ cells could not be compared since the only way to select for the $Lsh^{off/off}$ MEFs with geneticin treatment would have removed the converted cells from the population. However, that is the only possible explanation for the increase in the $Lsh^{off}$ allele and decrease in the converted allele when the cells are maintained in culture after the initial conversion to the $Lsh^{on}$

orientation. We also know from the previous rescue cell lines generated in the lab (*Lsh*$^{-/-}$ LSH MEFs) that *Lsh* knockout cells tend to become more sensitive and slow down their proliferation rate when LSH is re-introduced. This could explain why the *Lsh*$^{off/off}$ cells outgrow the converted cells after they start expressing LSH. Moreover, when the cells were treated with geneticin, a lower number of colonies was detected in the non-geneticin control dish for the OHT treated cells compared to the non-OHT treated cells. This result reinforces the idea of a lower proliferation rate of the *Lsh*$^{on/on}$ cells. The other issue concerned the decrease in the efficiency of conversion of the allele with higher passage numbers in the immortalised *Lsh*$^{off/off}$ MEFs. A recent study has shown that MCF7 cells began to lose their sensitivity to tamoxifen from the second passage. This is due to a decrease of drug influx via overexpression of the drug efflux transporters (Krisnamurti, Louisa, Anggraeni, & Wanandi, 2016). Interestingly, the *Lsh*$^{off/off}$ MEFs became resistant to tamoxifen without previous contact with the drug. Intrinsic resistance to tamoxifen could be acquired in the immortalised *Lsh*$^{off/off}$ MEFs, even without previous contact with tamoxifen (Saxena, Stephens, Pathak, & Rangarajan, 2011; Xia & Smith, 2012), but it is more likely that the ER-Cre virus is being silenced in the cells. The immortalisation of *Lsh*$^{off/off}$ MEFs using the retrovirus containing Large T antigen managed to circumvent the issue of the acquired resistance to tamoxifen. There is a higher percentage of cells with the allele in the *Lsh*$^{on}$ orientation until day fourteen of OHT treatment in Large T antigen immortalised cells. After this day the non-converted cells expand and outgrow the *Lsh*$^{on/on}$ population, generating a time frame of fourteen days to study the changes in chromatin produced by LSH that lead to epigenetic gene silencing and establishment of DNA methylation.

Several hypothesis exist for the function of LSH at repetitive genomic regions as well as developmentally-programed genes. One hypothesis for the role of LSH in DNA methylation is that LSH induces localised accumulation and association of chromatin modification factors after binding to a target site in the genome. LSH could play a role as a recruiting factor for DNMTs and HDACs to establish transcriptionally repressive chromatin. DNA methylation could be involved in either initiating or stabilizing any repressive chromatin state established at LSH targeted regions (Myant & Stancheva, 2008). Accordingly, studies analysing LSH involvement in pericentromeric heterochromatin formation showed association of LSH with repressive epigenetic regulators. LSH can for example associate with HP1, a heterochromatin regulator, (Yan, Cho et al. 2003) and this association could involve recruitment of DNMT3A or DNMT3B to the LSH target loci in the genome to promote DNA methylation (Zhu et al., 2006). Moreover, recent studies have shown that the role of LSH in the establishment of DNA methylation requires its ATPase activity (Ren et al., 2015; Termanis et al., 2016), suggesting a chromatin remodelling activity in addition to its role as an scaffolding protein (Myant &

Stancheva, 2008). As mentioned earlier, different nucleosome occupancy at repetitive elements has been shown in knockout ES cells when compared to wild-type ES cells after differentiation. This loss of nucleosome occupancy was restored to wild-type levels when wild-type LSH was re-expressed in the knockout cells but not in the ATP mutant LSH (Ren et al., 2015). Accordingly, it was suggested that the primary function of LSH was ATP-dependent nucleosome remodelling, which may promote *de novo* methylation in differentiating embryonic stem cells. In contrary to these studies, the nucleosome occupancy assay, NOMe assay, performed using chromatin from $Lsh^{off/off}$ and $Lsh^{+/+}$ MEFs did not detect a difference in nucleosome occupancy at the promoters of *Rhox* genes. The use of the conditionally reversible *Lsh* knockout system will allow to study whether LSH restoration involves nucleosome remodelling that promotes establishment of DNA methylation and whether this occurs at all LSH-dependent loci or only repetitive elements, as shown before (Ren et al., 2015). Moreover, re-expression of wild-type LSH in somatic knockout cells re-establishes DNA methylation and silencing of misregulated genes in a cell autonomous manner (Termanis et al., 2016). It would be interesting to know whether the previously suggested difference in nucleosome occupancy can also be rescued by LSH in non-differentiating cells. While other proteins from the SNF2-like subfamily have nucleosome remodelling activity (Hoffmeister et al., 2017; Oppikofer et al., 2017; Stockdale, Flaus, Ferreira, & Owen-hughes, 2006), ATP-dependent nucleosome remodelling activity by LSH had never been proven *in vitro* until a recent publication has demonstrated that LSH is part of a bipartite nucleosome remodelling complex. LSH alone cannot remodel nucleosomes, but the complex that LSH forms with CDCA7 possesses nucleosome remodelling activity. Furthermore, they show that CDCA7 is essential for binding of LSH to chromatin in cells (Jenness et al., 2018). Mutations in CDCA7, LSH, or DNMT3B cause immunodeficiency–centromeric instability–facial anomalies (ICF) syndrome (Thijssen et al., 2015). ICF syndrome is thought to be caused by defective DNA methylation although the molecular basis of the syndrome is unknown (G.-L. Xu et al., 1999). Importantly, the study concerning the LSH-CDCA7 complex also mentions that DNMT3A or DNMT3B were not detected on chromatin, suggesting a role for the nucleosome remodelling complex prior to and perhaps beyond DNA methylation.

To summarise, the results described and discussed in this chapter confirm that our conditionally reversible *Lsh* knockout cellular system is suitable for further experiments. This conditional knockout will allow us to re-express LSH in the $Lsh^{off/off}$ MEFs and study the time required for establishment of gene silencing and DNA methylation at LSH-dependent loci. Once this is determined, the understanding of the order of events and changes in chromatin modifications leading to gene silencing, will help to gain some insight into the specific role of LSH. The study of LSH-dependent mechanisms in gene silencing will be the aim of chapter 5.

# Chapter 5 - Mechanism of LSH dependent gene silencing

## 5.1 Introduction

The previous chapter validated the conditionally reversible *Lsh* knockout cellular system, making it suitable for further experiments. This conditional knockout allows us to re-express LSH in a controlled manner in the $Lsh^{off/off}$ MEFs and study the time required for acquisition of gene silencing and DNA methylation at LSH-dependent loci. Once the time required to rescue gene silencing in this *Lsh* knockout system is determined, the understanding of the order of events and changes in chromatin modifications leading to this silencing will help to gain further insight into the specific role of LSH.

Previous work carried out in the lab has shown that specific genes misregulated in $Lsh^{-/-}$ MEFs can be silenced and DNA methylation restored at their promoters when the wild-type LSH, but not the catalytically inactive mutant LSH, is introduced into these cells. LSH was able to silence genes that are marked at their promoter regions by active histone modifications such as H3K4me3 and H3Ac, and this silencing was accompanied by loss of the active chromatin marks. Importantly, LSH-dependent gene silencing occurred in somatic cells, in the absence of differentiation signals, suggesting that LSH functions in a cell autonomous manner (Termanis et al., 2016). It is not yet understood whether the changes in active histone marks are a secondary effect of the gene silencing or directly caused by the restoration of LSH expression in these cells. Other alterations in histone modifications have been reported in the absence of LSH, such as an increase of the histone modification H3K27me3 at promoter regions, suggesting potential redundancy of epigenetic silencing mechanisms (Yu et al., 2014). Moreover, LTR retrotransposons were shown to gain H3K4me3 in *Lsh* knockout cells and lose H3K9me3 at misregulated IAPs (Dunican et al., 2013). Interestingly, in reporter-based assays an indirect interaction of LSH with histone deacetylases (HDACs) via DNMT3B was shown to be necessary for LSH-dependent transcriptional repression, but this repression did not involve *de novo* DNA methylation (Myant & Stancheva, 2008). Another study has shown that the LSH-dependent repression of retrotransposons, such as IAPs, does not require the function of DNMT3B (Dunican et al., 2013). However, it has been suggested in many publications that the main role of LSH in epigenetic gene silencing is to facilitate DNA methylation by DNMT3B, and that chromatin remodelling by LSH is necessary for this function in DNA methylation (Ren et al., 2015). Accordingly, the homolog DDM1 in *Arabidopsis thaliana* promotes nucleosome repositioning in an ATP-dependent manner (Brzeski & Jerzmanowski, 2003) and is required for gene silencing and normal patterns of DNA methylation (J. Jeddeloh & Bender, 1998). The ATPase chromatin remodeling activity of LSH has been suggested

many times and a recent publication has shown that LSH, together with CDCA7, forms part of a nucleosome remodeling complex (Jenness et al., 2018).

In summary, the mechanism of LSH-dependent gene silencing is yet unclear. The use of the conditionally reversible *Lsh* knockout system allows the changes in chromatin modifications leading to gene silencing to be examined in more detail. Therefore such studies will help to determine the order of events leading to gene silencing after restoration of LSH expression and gain clearer insights into the role of LSH in epigenetic gene silencing.

## 5.2 LSH expression leads to silencing of misregulated genes in *Lsh*$^{off/off}$ MEFs

As mentioned earlier, when wild-type LSH is re-expressed in *Lsh* knockout MEFs, the gain of transcription of specific protein coding genes and repetitive elements in the absence of LSH can be rescued to the initial silenced state in a cell autonomous manner (Termanis et al., 2016). The *Lsh*$^{off/off}$ MEFs used in this study also show gain of expression at specific genes and repetitive elements as shown in the previous chapter. To determine whether re-expression of LSH in *Lsh*$^{off/off}$ MEFs can also rescue the silenced state of these misregulated genes, *Lsh*$^{off/off}$ cells were treated with OHT and gene expression analysed by qRT-PCR. Specific protein coding genes including *Gm9, Rhox2, Rhox6* and *Tnfsf9* were studied, as well as IAP retrotransposons. *Gm9, Rhox2* and *Rhox6* genes were chosen from all the LSH-dependent loci since we knew that they were re-silenced when LSH was reintroduced in *Lsh* knockout MEFs. *Tnfsf9* was chosen because LSH re-expression in knockout cells did not rescue DNA methylation at *Tnfsf9* promoter but we were not aware of its expression state in these cells. It was interesting to know whether the gain of expression of *Tnfsf9* was re-silenced after LSH expression even though DNA methylation was not rescued at its promoter.

The *Lsh*$^{off/off}$ MEFs were treated with OHT and samples collected after LSH re-expression (see Figure 4.14), at day seven and day fourteen of treatment, to study whether there was gene silencing in these cells. The protein coding genes *Gm9, Rhox2* and *Rhox6* showed gain of transcription in the absence of LSH and they were silenced after fourteen days of OHT treatment (Figure 5.1 A). Surprisingly, *Tnfsf9* did not gain transcription in *Lsh*$^{off/off}$ cells compared to *Lsh*$^{+/+}$ MEFs, and acquired higher expression after seven and fourteen days of OHT treatment. Interestingly, some experiments showed gain of transcription of *Tnfsf9* in the *Lsh*$^{off/off}$ MEFs, but *Tnfsf9* was never silenced after OHT treatment. The gain of expression of IAP retrotransposons in *Lsh*$^{off/off}$ cells was also rescued to a silent state after OHT treatment. LSH expression was restored to wild-type level after seven days of OHT treatment as shown in the previous chapter. Accordingly, at day seven of OHT treatment there was a detectable decrease in the expression of all examined loci that are susceptible to re-silencing in the

presence of LSH, except *Rhox6,* which showed a transient increase in the expression. At day fourteen of OHT treatment all these loci were silenced although they showed different degrees of downregulation. It is important to mention that there was a lot variation between experiments since the non-converted $Lsh^{off/off}$ cells were outgrowing at different rates the $Lsh^{on/on}$ MEFs. This caused big differences in the gene expression after OHT treatment, depending on the number of cells with the allele in the non-converted or converted $Lsh^{on}$ orientation. The same experiment was performed, but the cells were treated with DMSO as a control. This control was done to confirm that silencing of misregulated genes in $Lsh^{off/off}$ MEFs was caused by the conversion of the allele to $Lsh^{on}$ orientation and LSH expression upon OHT treatment, and was not a secondary effect of the treatment or DMSO used as a solvent. The control experiment demonstrated that DMSO treatment did not cause silencing of any of the loci in study (Figure 5.1 B). In conclusion, LSH expression after OHT treatment of $Lsh^{off/off}$ MEFs leads to silencing of misregulated genes such as *Gm9, Rhox2, Rhox6,* and IAP retrotransposons in the converted cells.

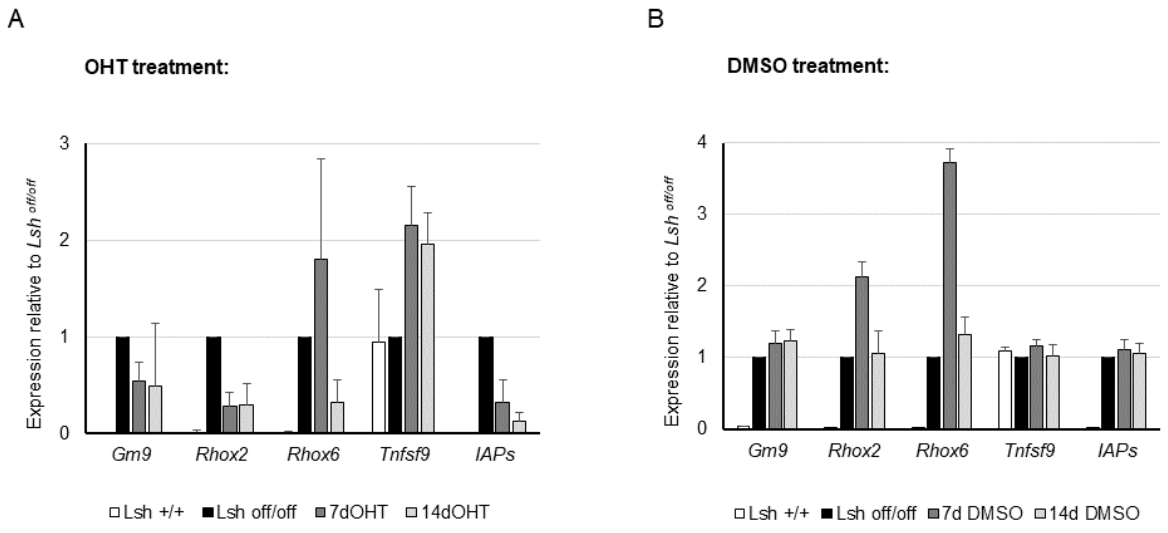


**Figure 5.1. LSH expression after OHT treatment leads to silencing of misregulated genes in $Lsh^{off/off}$ MEFs.**

Expression of IAP retrotransposons and LSH-dependent genes in $Lsh^{+/+}$ MEFs, $Lsh^{off/off}$ MEFs and $Lsh^{off/off}$ MEFs after OHT (A) and DMSO (B) treatment for a period of seven and fourteen days. The graphs show the change in expression relative to $Lsh^{off/off}$ MEFs as assessed by quantitative RT-PCR. The housekeeping *Gapdh* gene was used for normalisation. The error bars represent the standard deviation.

### 5.3 Silencing of LSH-dependent genes occurs without gain of DNA methylation

Specific genes and repetitive elements that are misregulated in *Lsh*$^{-/-}$ MEFs, also lose DNA methylation from their promoters (Myant et al., 2011). We have shown in the previous chapter that the misregulation of LSH-dependent loci in *Lsh*$^{off/off}$ MEFs is also accompanied by loss of DNA methylation (see Figure 4.4). The previous section of this chapter has demonstrated that LSH re-expression in *Lsh*$^{off/off}$ MEFs can silence these loci. Since LSH can restore DNA methylation when expressed in long-term in the null MEFs (Termanis et al., 2016), I next examined whether or not DNA methylation was re-established simultaneously with gene silencing. I also wanted to determine the time required to restore DNA methylation at the promoters of protein coding genes and IAP retrotransposons in *Lsh*$^{off/off}$ MEFs, to better understand the order of events leading to gene silencing after OHT treatment.

*Lsh*$^{off/off}$ MEFs were treated with OHT and samples collected after LSH is expressed to determine whether DNA methylation was rescued simultaneously with or after gene silencing. As previously, the expression of *Gm9* gene and IAP retrotransposons was examined at days seven and day fourteen of OHT treatment in *Lsh*$^{off/off}$ primary MEFs. DNA methylation at the promoter of *Gm9* and LTR-embedded promoter of IAPs was examined at the same time points, day seven and fourteen, by bisulfite sequencing. As expected, the gain of expression observed for *Gm9* gene in *Lsh*$^{off/off}$ MEFs was downregulated after seven days of OHT treatment and the gene was completely silenced after fourteen days (Figure 5.2 A). IAPs expression started to be downregulated after seven days of OHT treatment and was completely repressed at day fourteen (Figure 5.2 B). The silencing of both loci in primary MEFs at day fourteen of treatment is in agreement with the previously obtained results from immortalised MEFs (Figure 5.1 A). There was an overall loss of DNA methylation by 36% from promoter of *Gm9* and 34% from LTR of IAPs in *Lsh*$^{off/off}$ MEFs compared to *Lsh*$^{+/+}$ MEFs (Figure 5.2 C and D, respectively). Surprisingly, the percentage of DNA methylation at the two time points after OHT treatment was similar to the *Lsh*$^{off/off}$ MEFs, showing no rescue of DNA methylation at these promoters when LSH was expressed in the cells and genes silenced (Figure 5.2). DNA methylation at the promoter of *Gm9* decreased after seven days of treatment and then slightly increased at day fourteen compared to *Lsh*$^{off/off}$ MEFs (Figure 5.2 C). On the contrary, DNA methylation at IAPs decreased after fourteen days of treatment compared to *Lsh*$^{off/off}$ cells (Figure 5.2 D). However, the variation in the percentages of DNA methylation was small and could be due to the limited number of sequenced clones for each time point. Taken together, these results show that LSH-dependent loci are silenced after fourteen days of OHT treatment and this happens without significant gain of DNA methylation at their promoters in primary MEFs.

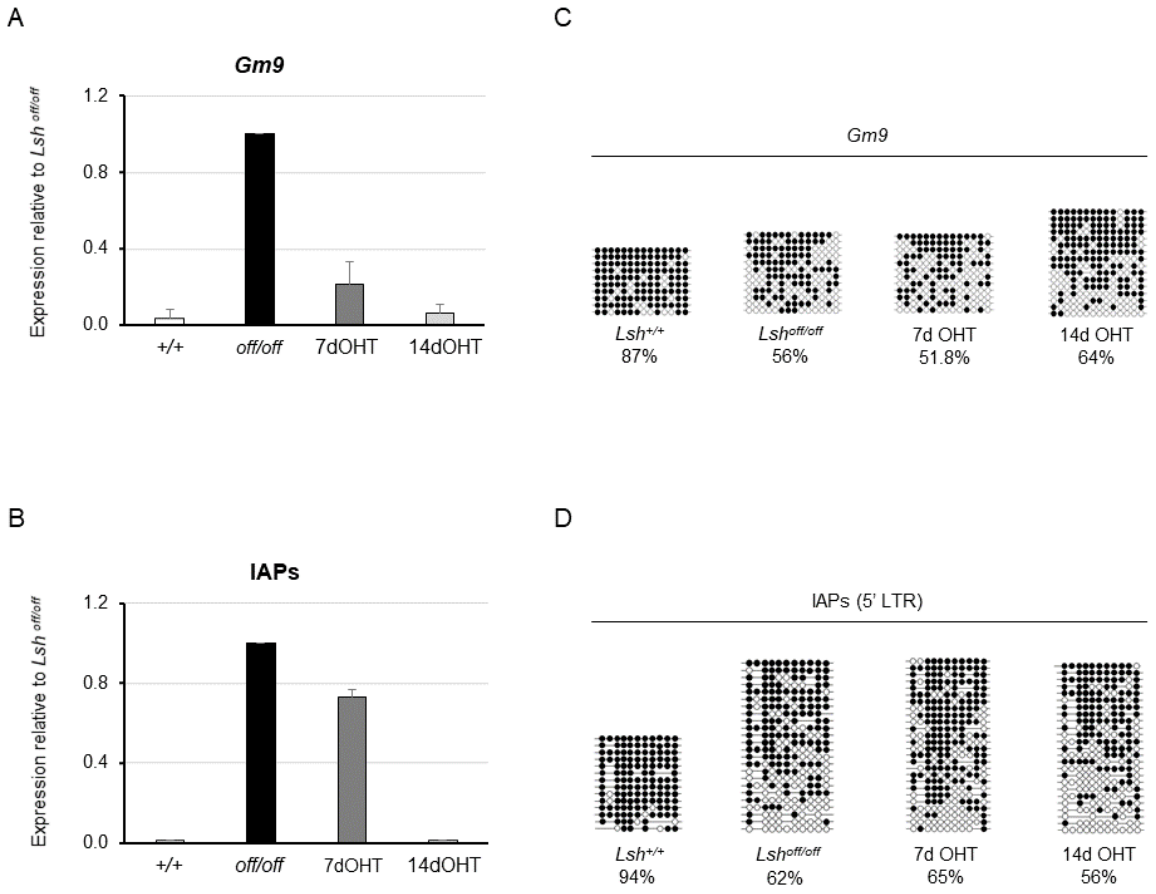


**Figure 5.2. Silencing of *Gm9* and IAPs upon LSH restoration occurs without rescue of DNA methylation at their promoters in primary MEFs.**

Expression of *Gm9* (A) and IAPs (B) in $Lsh^{+/+}$ MEFs, $Lsh^{off/off}$ MEFs and $Lsh^{off/off}$ MEFs after OHT treatment for a period of seven and fourteen days. The graphs show the change in expression relative to $Lsh^{off/off}$ MEFs as assessed by quantitative RT-PCR. The housekeeping *Gapdh* gene was used for normalisation. The error bars represent the standard error of the mean calculated from three technical replicates. Sodium bisulfite DNA sequencing for *Gm9* (C) and IAPs (D) promoters in $Lsh^{+/+}$ MEFs, $Lsh^{off/off}$ MEFs and $Lsh^{off/off}$ MEFs after OHT treatment. Each row in the diagrams represents a single DNA strand and methylated CpGs are shown as black circles. The percentage of methylation is indicated below the diagrams.

This analysis was done in primary MEFs, but since we observed differences in DNA methylation between primary and immortalised cells, I wanted to determine whether there was a difference in the rescue of DNA methylation in immortalised cells undergoing OHT treatment and gene silencing. DNA methylation at LTR-embedded promoter of IAPs was analysed by bisulfite sequencing using gDNA from LTag immortalised MEFs. There was an overall 62% loss of DNA methylation in *Lsh*$^{off/off}$ MEFs compared to *Lsh*$^{+/+}$ MEFs (Figure 5.3). The loss of DNA methylation was higher than the 34% observed in primary cells (Figure 5.2 D), which was in agreement with the observation mentioned in chapter 4. The DNA methylation at IAPs at day fourteen of OHT treatment was 50% in LTag immortalised MEFs (Figure 5.3). This result indicated a mild increase in the methylation after OHT treatment at immortalised cells although the overall percentage of DNA methylation was similar to the 56% value obtained for primary MEFs. This increase in DNA methylation was similar to the increase observed at *Gm9* promoter in primary MEFs and could be also due to the limited number of sequenced clones. There was still a 46% difference in DNA methylation at IAPs between *Lsh*$^{+/+}$ cells and fourteen days OHT-treated MEFs, which was similar to previous values obtained for *Lsh* knockout cells.

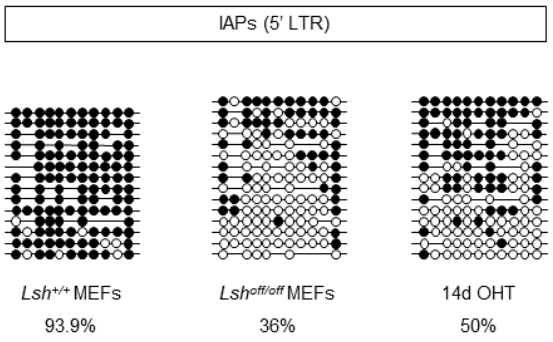


**Figure 5.3. There is no rescue of DNA methylation at IAPs upon LSH restoration in LTag immortalised MEFs.**

Sodium bisulfite DNA sequencing for IAPs (5' LTR) in *Lsh*$^{+/+}$ MEFs, *Lsh*$^{off/off}$ MEFs and *Lsh*$^{off/off}$ MEFs after fourteen days of OHT treatment. Each row in the diagrams represents a single DNA strand and methylated CpGs are shown as black circles. The percentage of methylation is indicated below the diagrams.

In conclusion, silencing of LSH-dependent loci after OHT treatment occurs without rescue of DNA methylation at their promoters in primary and immortalised MEFs. Therefore this indicates that promoting DNA methylation might not be the primary function of LSH in epigenetic gene-silencing as previously hypothesized.

## 5.4 LSH expression leads to changes in histone modifications and histone density

Previously we have shown that histone modifications, such as H3K4me3 and H3Ac, associated with active transcription and found at genes that are misregulated in *Lsh* knockout cells are loss when LSH is reintroduced in these cells and genes re-silenced. This re-silencing of misregulated genes is accompanied by gain of DNA methylation at the promoters of the genes (Termanis et al., 2016) upon sustained expression of LSH. Surprisingly, when $Lsh^{off/off}$ MEFs are treated with OHT and LSH expression is restored in these cells, there is silencing of LSH-dependent loci after fourteen days of treatment but not gain of DNA methylation during this period. To better understand the role of LSH in gene silencing, I analysed whether changes in histone modifications that may lead to silencing occur after OHT treatment. Moreover, other studies have shown that LSH, when re-expressed in differentiating *Lsh* knockout ES cells, can rescue normal patterns of nucleosome density at repetitive sequences (Ren et al., 2015). Therefore, I also analysed histone density at the promoters of misregulated genes after OHT treatment of the $Lsh^{off/off}$ cells to determine whether LSH plays a role as a chromatin remodeler in the conditionally reversible *Lsh* knockout system.

Before attributing the possible changes in chromatin structure and modifications to LSH, I first determined whether LSH localises at the loci of interest that were silenced after OHT treatment in $Lsh^{off/off}$ MEFs. Previously, the lack of appropriate commercial antibodies against LSH that could be used in ChIP experiments had made it difficult to study the localization of LSH to specific loci, but luckily a new commercially available polyclonal antibody enabled me to perform such ChIP experiments. As mentioned earlier, LSH was detected after treating the $Lsh^{off/off}$ MEFs for four days with OHT but its expression was not restored to wild-type level until day seven of OHT treatment when assessed by Western blot. ChIP experiments for LSH were carried out using crosslinked chromatin from $Lsh^{+/+}$ MEFs, $Lsh^{off/off}$ MEFs and $Lsh^{off/off}$ cells treated for seven and fourteen days with tamoxifen. The amount of LSH detected in $Lsh^{off/off}$ MEFs was considered as a background signal produced by the antibody at each locus. The ChIP experiments showed that LSH binds at LSH-dependent loci in $Lsh^{+/+}$ MEFs, which was somewhat unexpected given that LSH displays a very dynamic chromatin binding behaviour (Lungu et al., 2015). Interestingly, I detected LSH at most examined loci, including the control promoters of *Actin* and *Tnfsf9* genes, after seven days of OHT treatment (Figure 5.4 A). However, LSH was not detected at *Rhox2* gene after OHT treatment. The location of LSH at the promoters of these loci matched the time required for the protein levels of LSH to be restored to wild-type levels after OHT treatment. After fourteen days of OHT treatment there was a loss of LSH from all examined loci with the exception of *Rhox2* (Figure 5.4 A). This observation could be explained by the dynamic association of LSH with chromatin, as mentioned earlier. Once I detected LSH at the loci of interest, I examined whether or not

changes in histone density occur after LSH restoration in the *Lsh*$^{off/off}$ MEFs. ChIP for histone H3 was performed using the control cell lines and the same time points of OHT treatment in *Lsh*$^{off/off}$ MEFs (Figure 5.4 B). The occupancy of H3 at the promoters of *Actin, Gm9* and *Rhox6* was higher in *Lsh*$^{off/off}$ cells than *Lsh*$^{+/+}$ MEFs. However, *Rhox2, Tnfsf9* and IAPs showed similar H3 occupancy in both cell lines. There was a general decrease in H3 occupancy after seven days of tamoxifen treatment at the promoters of most studied loci, which was in agreement with the presence of LSH at these promoters at the same time point. However, after fourteen days of OHT treatment the examined loci did not follow a similar trend. Some promoters maintained a lower histone occupancy while others showed an increase in H3 occupancy. Moreover, there was no correlation between the loci that showed presence of LSH at their promoters at day fourteen of treatment and the loci maintaining a lower H3 occupancy at the same time point. Taken together, these results demonstrate that LSH localises to LSH-dependent loci as well as the control loci *Actin* and *Tnfsf9*, and leads to a mild decrease in H3 occupancy after seven days of tamoxifen treatment. However, the change in histone density is not maintained in all the studied loci after fourteen days of treatment.

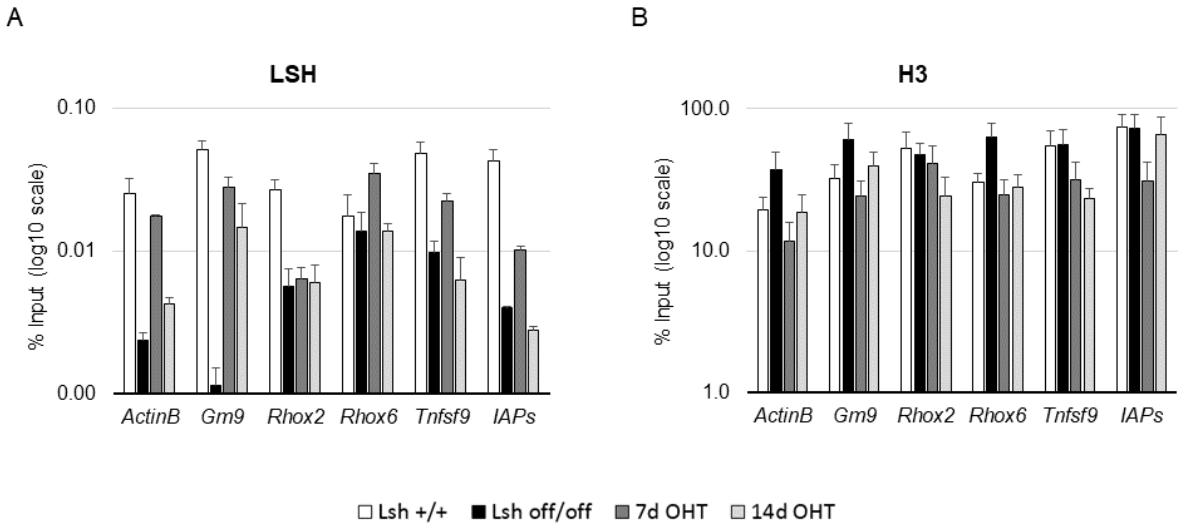


**Figure 5.4. LSH localises at LSH-dependent loci and leads to a mild decrease in H3 occupancy after 7 days of tamoxifen treatment at the promoter region of specific protein coding genes and IAPs.**

ChIP for LSH (A) and histone H3 (B) in *Lsh*$^{+/+}$ MEFs, *Lsh*$^{off/off}$ MEFs and *Lsh*$^{off/off}$ MEFs after OHT treatment. *Actin* was used as a control since its transcriptional state is not expected to change in the cells in the study. *Tnfsf9* was used as a negative control since it is not silenced after OHT treatment when LSH is expressed in the *Lsh*$^{off/off}$ MEFs. The error bars represent standard error of the mean calculated from two independent experiments.

To determine whether LSH restoration in *Lsh*$^{off/off}$ MEFs, which leads to a decrease in H3 occupancy at the promoters of LSH-dependent loci, also facilitates changes in histone modifications, I carried out ChIP experiments to examine H3Ac and H3K9me2 modifications. As expected, the *Lsh*$^{+/+}$ cells were marked by H3Ac at the promoter of the control gene *Actin* compared to the non-expressed LSH-dependent loci where we barely observed this mark (Figure 5.5 A). There was enrichment for the H3Ac modification at the promoters of all the loci in study in *Lsh*$^{off/off}$ cells compared to *Lsh*$^{+/+}$ MEFs. The enrichment was higher for the loci that gained transcription in the *Lsh*$^{off/off}$ MEFs, as expected since H3Ac mark is associated with active transcription. After seven days of OHT treatment there was loss of H3Ac mark from the promoters of LSH-dependent genes as well as the controls *Actin* and *Tnfsf9,* reaching similar to wild-type levels of these modifications for each locus. The loss of H3Ac mark occurred at simultaneously with LSH binding at these promoters and the decrease in H3 occupancy. However, this loss of H3Ac mark was not maintained in LSH-dependent loci at day fourteen of treatment, showing a mild increase that did not reach the levels in *Lsh*$^{off/off}$ MEFs. Contrary to this, the level of H3Ac mark at the controls *Actin* and *Tnfsf9* was restored to the *Lsh*$^{off/off}$ MEFs levels (Figure 5.5 A). The expression of both control genes did not change in *Lsh*$^{off/off}$ MEFs and cells treated for fourteen days with OHT, which might explain the similar levels of H3Ac mark at their promoters. The gain of H3Ac at LSH-dependent loci at day fourteen of OHT treatment could be explained by the expansion of non-converted cells which at this point start outgrowing the *Lsh*$^{on/on}$ MEFs.

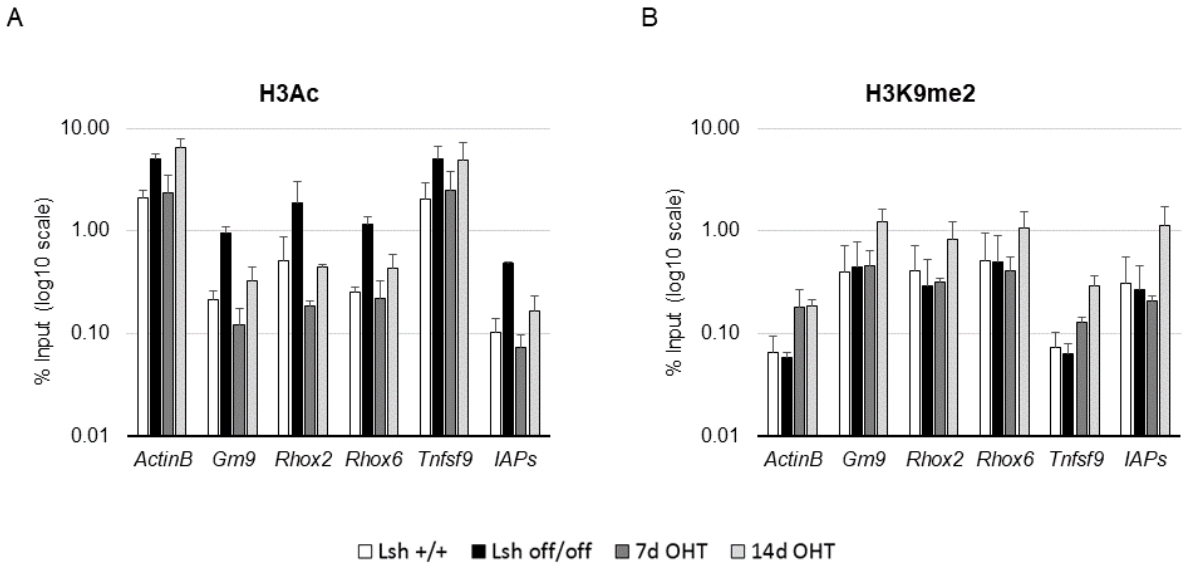


**Figure 5.5. There is loss of H3Ac and gain of H3K9me2 modification at the promoters of LSH-dependent loci upon LSH expression after OHT treatment.**

ChIP for H3Ac modification (A) and histone H3K9me2 modification (B) in *Lsh*$^{+/+}$ MEFs, *Lsh*$^{off/off}$ MEFs and *Lsh*$^{off/off}$ MEFs after seven and fourteen days of OHT treatment. *Actin* was used as a control since its transcriptional state does not change in the cells in study. *Tnfsf9* was used as a negative control since it is not silenced after OHT treatment when expressed in *Lsh*$^{off/off}$ MEFs. The error bars represent the standard error of the mean calculated from two independent experiments.

Once I detected loss of H3Ac mark from LSH-dependent loci after seven days of OHT treatment, I decided to determine whether histone modification H3K9me2 has a role in LSH-dependent gene silencing in the absence of DNA methylation. I decided to study this histone mark since LSH has been previously shown to cooperate with the G9a/GLP complex, that catalyses mono and di-methylation of H3K9, to generate normal DNA methylation patterns and stable gene silencing at specific promoters (Myant et al., 2011). I performed ChIP for H3K9me2 with the same cell lines and time points used in previous ChIP experiments. The level of H3K9me2 mark was similar in *Lsh*$^{off/off}$ and *Lsh*$^{+/+}$ MEFs at all examined loci, although some of the specific genes and IAPs were only expressed in the knockout cells (Figure 5.5 B). After seven days of OHT treatment there was a small increase of H3K9me2 mark at the promoters of the controls *Actin* and *Tnfsf9*, but not at LSH-dependent loci. Interestingly, at day fourteen of the treatment there was a substantial gain of H3K9me2 mark at the promoters of all LSH-dependent loci, including *Tnfsf9* gene. However, the gain of H3K9me2 mark at day fourteen was higher at LSH-dependent loci than at the control genes *Actin* and *Tnfsf9.* This indicates that the G9a/GLP complex might be facilitating gene silencing upon LSH restoration. The differential H3 occupancy (Figure 5.4 B) has potential implications on the results shown here since the ChIP data has not been normalized to H3 occupancy. The loss of H3Ac after

seven days of treatment would be less obvious due to the lower occupancy of H3 at this time point. However, the gain of H3K9me2 mark at day fourteen of treatment would increase if the data was normalised to H3.

In conclusion, LSH localises to the promoters of LSH-dependent loci upon restoration of its expression following the OHT treatment, and leads to a decrease in H3 occupancy at the promoters of specific genes. The transcriptional silencing of these genes and repetitive elements is accompanied by loss of H3Ac and subsequent gain of H3K9me2 at their promoter region.

### 5.5 The role of HDACs and G9a/GLP in LSH-dependent gene silencing

Since the first changes in histone modifications observed upon restoration of LSH expression in the conditionally reversible *Lsh* knockout cells are the loss of H3Ac and subsequent gain of H3K9me2 at the promoters of LSH-dependent loci, I decided to investigate the role of HDACs and the G9a/GLP complex in LSH-dependent gene silencing. More specifically, I wanted to determine whether the activity of HDACs and G9a/GLP enzymes is required for the transcriptional repression induced by LSH. I used two drugs for this purpose, valproic acid (VPA) and UNC 0638 (UNC), which are potent inhibitors of HDACs and G9a/GLP, respectively. Valproic acid was used since it has been shown that cells tolerate better this drug than any other HDAC inhibitor (Gottlicher et al., 2001).

To clarify whether removal of H3Ac by HDACs is necessary for LSH-dependent gene silencing, $Lsh^{off/off}$ MEFs were treated with OHT to convert the cassette to the $Lsh^{on}$ orientation and then VPA was added to the cell culture medium after four days of OHT treatment when LSH is starting to be expressed (Figure 5.6 A). Samples were collected at the beginning of the experiment and after seven and fourteen days of treatment. The efficiency of the HDACs inhibitor was assessed by Western blot using antibodies against histone H3 acetylated at K9 and K14 and histone H4 as a control. Samples from $Lsh^{off/off}$ MEFs treated with OHT and UNC were also included in this experiment as an extra control since we were not expecting changes in the levels of H3Ac in these samples. An increase in H3Ac mark was expected for the samples treated with OHT and VPA compared to the control samples, only OHT and OHT plus UNC treatment. However, there was no detectable increase in H3Ac mark in the samples treated with VPA, especially when comparing to the OHT-treated samples (Figure 5.6 B).This suggests that the drug was not working or that the converted OHT-treated cells were dying under VPA treatment. To verify that the VPA treatment was not working in the OHT-treated MEFs and it was not an issue with the concentration or the drug itself, a control experiment was carried out. $Lsh^{+/+}$ and $Lsh^{off/off}$ MEFs were treated with VPA and sodium butyrate (NaB),

which is another widely used inhibitor of HDACs. The concentration of VPA in the control experiment was the same as the one used previously and as described by others (Gottlicher et al., 2001). The treatment of *Lsh*$^{+/+}$ and *Lsh*$^{off/off}$ MEFs with both inhibitors, VPA and NaB, caused an increase in H3Ac compared to the non-treated control samples as assessed by Western blot (Figure 5.6 C). The quantification of the Western blot showed the fold change in H3Ac after VPA and NaB treatment. The fold change increase in H3Ac relative to H4 was higher for NaB than VPA treatment in both cell lines, which indicates that the treatment with NaB was more efficient than the VPA treatment (Figure 5.6 D). The difference in the treatment of *Lsh*$^{off/off}$ MEFs with only VPA and VPA plus OHT could indicate a higher sensitivity of the OHT-converted cells to VPA. This would explain why there is no increase in H3Ac mark when the cells are treated at the same time with OHT and VPA, since the converted cells might not tolerate the inhibition of HDACs. The inefficient VPA treatment combined with OHT treatment in the *Lsh*$^{off/off}$ MEFs, did not allow us to study the role of HDACs in LSH-dependent gene silencing.

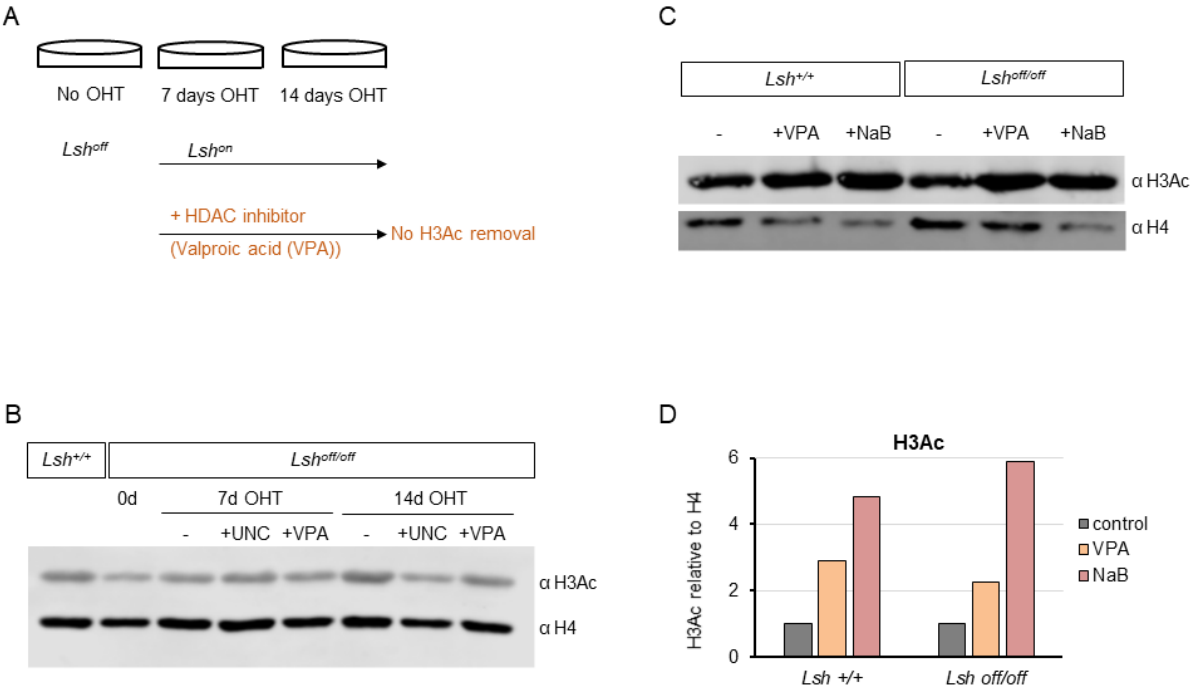


**Figure 5.6. There is no change in the levels of H3Ac in *Lsh*$^{off/off}$ MEFs after treatment with OHT and VPA, although VPA treatment leads to an increase in H3Ac in *Lsh*$^{+/+}$ MEFs and *Lsh*$^{off/off}$ MEFs.**

A) Schematic diagram of the sequential treatment of *Lsh*$^{off/off}$ MEFs with OHT and VPA. B) Western blot for H3Ac modification with an antibody to detecting acetylation of H3 at K9/K14. *Lsh*$^{off/off}$ MEFs were treated with OHT and VPA in addition to OHT (+VPA). Samples were collected after seven days and fourteen days of treatment to analyse the histone modifications. *Lsh*$^{+/+}$ and *Lsh*$^{off/off}$ 0d were used as controls for the initial H3Ac levels. +UNC samples are used as an additional control. H4 is a loading control. C) Western blot detecting H3Ac levels in *Lsh*$^{+/+}$ and *Lsh*$^{off/off}$ MEFs treated with VPA and NaB for a period of three days. Samples from untreated *Lsh*$^{+/+}$ and *Lsh*$^{off/off}$ MEFs were used to detect the initial H3Ac levels. H4 is a loading control. D) Quantification of H3Ac in the Western blot shown in (C). H3Ac intensity relative to H4 for *Lsh*$^{+/+}$ and *Lsh*$^{off/off}$ MEFs was normalised to the control non-treated samples.

I also treated the *Lsh*$^{off/off}$ MEFs with OHT and UNC to determine whether deposition of H3K9me2 at promoters of LSH-dependent loci by G9a/GLP is required for gene silencing. The sequential treatment with OHT and UNC (Figure 5.7 A) was performed as the VPA treatment described earlier. A decrease in the global levels of H3K9me2 modification was expected after UNC treatment. This was initially assessed by Western blot, including non-treated *Lsh*$^{+/+}$ and *Lsh*$^{off/off}$ MEFs as controls together with samples from *Lsh*$^{off/off}$ MEFs treated only with OHT and with OHT plus UNC for seven and fourteen days. The OHT and VPA treated samples were also included as an additional control. There was no clear difference in the global levels of H3K9me2 after seven days of treatment (Figure 5.7 B). However, after fourteen days of combined OHT with UNC treatment there was a decrease in the global levels of H3K9me2

compared to the control samples. To verify this globally observed difference in a locus specific manner, ChIP for H3K9me2 was carried out for LSH-dependent loci. The ChIP was performed with non-treated control samples from *Lsh*$^{+/+}$ and *Lsh*$^{off/off}$ MEFs, and samples from *Lsh*$^{off/off}$ MEFs treated only with OHT and OHT with UNC for seven and fourteen days. H3K9me2 mark was barely detectable at the promoters of the control genes, *Actin* and *Tnfsf9*, which is in agreement with their unaltered expression during the OHT treatment (Figure 5.7 C). There were no clear differences in H3K9me2 mark between the OHT-treated and OHT with UNC-treated samples after seven days of treatment, as observed earlier in the Western blot. However, at day fourteen of OHT treatment there was a big gain of H3K9me2 mark at the OHT-treated samples but no gain of H3K9me2 was observed at the OHT with UNC-treated samples. This demonstrates that the UNC inhibitor produces the expected effect when applied to the OHT-converted *Lsh*$^{off/off}$ MEFs for a period of fourteen days. It also demonstrates that either G9a or GLP are involved in the LSH-dependent deposition of H3K9me2.

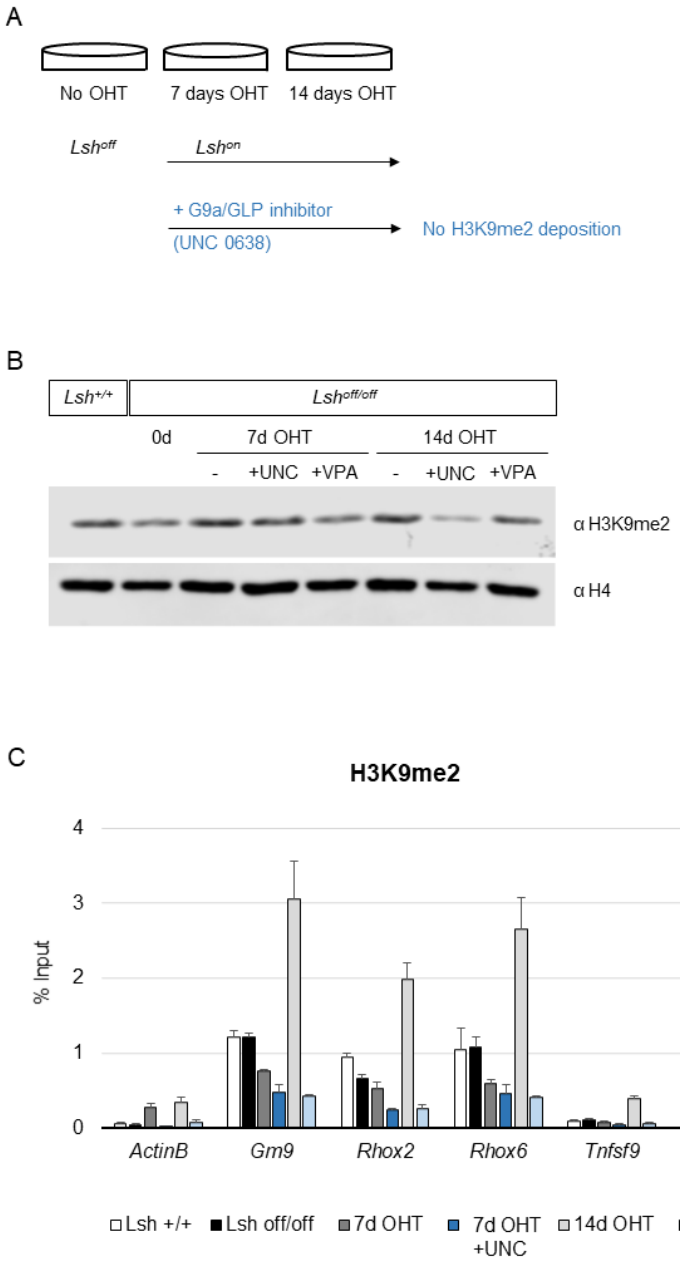


**Figure 5.7. There is no deposition of H3K9me2 mark at LSH-dependent loci after treatment of *Lsh*$^{off/off}$ MEFs with OHT and UNC, an inhibitor of G9a/GLP.**

A) Schematic diagram of the sequential treatment of *Lsh*$^{off/off}$ MEFs with OHT and UNC. B) Western blot for H3K9me2 modification in *Lsh*$^{off/off}$ MEFs treated with OHT and UNC in addition to OHT (+UNC). Samples were collected after seven days and fourteen days of treatment to analyse the histone modification after the different conditions. *Lsh*$^{+/+}$ and *Lsh*$^{off/off}$ 0d were used as controls for the initial H3K9me2 levels. +VPA samples are used as an additional control. H4 is a loading control. C) ChIP for H3K9me2 in *Lsh*$^{+/+}$ MEFs, *Lsh*$^{off/off}$ MEFs and *Lsh*$^{off/off}$ MEFs after OHT and OHT +UNC treatment. *Actin* was used as a control since its transcription does not change in the cells under study. *Tnfsf9* was used as a negative control since it is not silenced after OHT treatment of the *Lsh*$^{off/off}$ MEFs. The error bars represent standard error of the mean calculated from three technical replicates.

To determine further whether the inhibition of G9a/GLP activity has an effect on LSH-dependent gene silencing, I analysed the expression of specific protein coding genes and IAP retrotransposons by quantitative RT-PCR. *Rhox2* and *Rhox6* genes were silenced after fourteen days of OHT treatment, as expected, but this silencing was not observed when the cells were treated with OHT together with UNC for the same period (Figure 5.8 A). However, there were no clear differences in the expression of *Rhox* genes at day seven between the OHT and OHT plus UNC treated samples, which correlates with the effect of the inhibitor on H3K9me2 at day seven. In this experiment *Gm9* gene was not highly expressed in the $Lsh^{off/off}$ MEFs compared to $Lsh^{+/+}$ MEFs and this weak gain of expression was not affected by either the OHT or the OHT with UNC treatment. Similarly, there was no gain of expression of *Tnfsf9* in the $Lsh^{off/off}$ MEFs and therefore there were no changes observed after the different treatments. Surprisingly, the expression of IAPs in the $Lsh^{off/off}$ MEFs was reduced in both cases, in the OHT and the OHT plus UNC treated samples. After seven days of treatment we observed the same decrease in IAPs expression for both conditions (Figure 5.8 B) and at day fourteen of treatment IAPs were silenced in both samples. It seems that, in contrast to the *Rhox* genes, IAPs do not require deposition of H3K9me2 by the G9a/GLP complex for silencing upon LSH expression. To understand the silencing of IAPs after inhibition of G9a/GLP activity, I performed ChIP for H3K27me3 since this histone modification has been shown to play a role in IAPs silencing in the absence of DNA methylation (Walter, Teissandier, Pérez-Palacios, & Bourc'his, 2016). The changes in this histone mark were also analysed for the control gene *Actin* and IAPs in the control cell lines and $Lsh^{off/off}$ MEFs after seven and fourteen days after different treatments, OHT and OHT with UNC. $Lsh^{off/off}$ MEFs were marked by H3K27me3 at promoters of IAPs compared to $Lsh^{+/+}$ MEFs where we did not observe this mark (Figure 5.8 C). This gain of H3K27me3 in $Lsh^{off/off}$ cells could be responsible for the silencing of some IAPs in the absence of LSH. After seven days of OHT treatment, there was further gain of H3K27me3 which could be responsible for the silencing of IAPs when LSH expression was restored. The gain of H3K27me3 was higher in the OHT with UNC treated cells, which could be compensating for the lack of H3K9me2 deposition when the action of G9a/GLP enzymes is impaired. However, at day fourteen of treatment, there was loss of H3K27me3 in both samples, the OHT and OHT with UNC treated cells. This indicates that the role of H3K27me3 mark in the silencing of IAPs is transient in the absence of G9a/GLP activity when LSH is restored and suggests that another mechanism must be playing a role to maintain IAPs repressed after fourteen days of treatment. The gain of H3K27me3 between $Lsh^{off/off}$ cells and seven days of OHT and UNC treated cells can justify the level of repression of IAPs. The limited gain of the repressive mark H3K27me3 could be explained by limiting Polycomb Complex. However, this mark could be triggering other mechanisms to continue the silenced expression state, such as ubiquitylation of H2A by the PRC1 complex (Boyer et al., 2006).

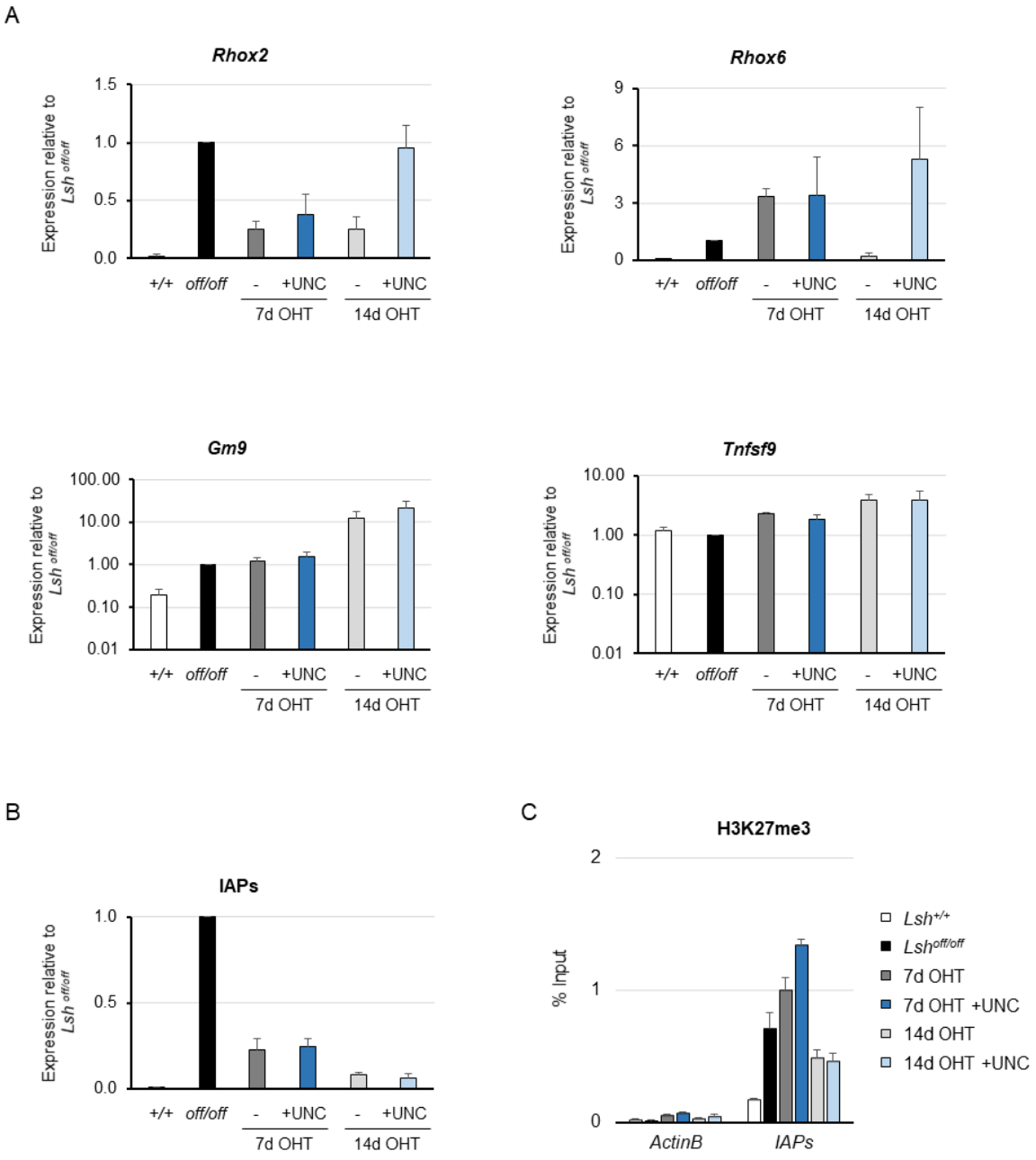


**Figure 5.8. H3K9me2 deposition by G9a/GLP is required for LSH-dependent gene silencing of *Rhox* genes while H3K27me3 seems to be playing a transitory role in the silencing of IAP retrotransposons in the absence of G9a/GLP activity.**

Expression of specific protein coding genes (A) and IAP retrotransposons (B) in *Lsh*$^{+/+}$, *Lsh*$^{off/off}$ and *Lsh*$^{off/off}$ MEFs after OHT and OHT with UNC (+UNC) treatment for a period of seven and fourteen days. The graphs show the change in expression relative to *Lsh*$^{off/off}$ MEFs assessed by quantitative RT-PCR. The housekeeping *Gapdh* gene was used for normalisation. The error bars represent standard error of the mean from two independent experiments. C) ChIP for H3K27me3 at IAPs in *Lsh*$^{+/+}$, *Lsh*$^{off/off}$ and *Lsh*$^{off/off}$ MEFs after OHT and OHT with UNC treatment. *Actin* was used as a control since its transcriptional state is not expected to change upon different treatments. The error bars represent standard error of the mean calculated from three technical replicates.

To summarise, the inhibition of G9a/GLP impairs the silencing of *Rhox* genes upon restoration of LSH. However, in the absence of H3K9me2 deposition by the G9a/GLP complex, the IAP retrotransposons are silenced and H3K27me3 seems to play a transient role in this silencing. To investigate whether there are other changes in histone modifications in the absence of G9a/GLP activity that could contribute to gene silencing, I performed ChIP for H3K9me3. At day fourteen of OHT treatment there was gain of H3K9me3 at the promoters of the protein coding genes, *Gm9*, *Rhox2* and *Rhox6*, which accompanied the silencing of *Rhox* genes mentioned earlier (Figure 5.9 A). However, this gain of H3K9me3 was not observed when the cells were treated with OHT and UNC. Accordingly, IAPs only gained H3K9me3 mark at their promoters after fourteen days of OHT treatment. Importantly, the gain of expression of IAPs in $Lsh^{off/off}$ MEFs was accompanied by reduction of H3K9me3 at their promoter regions compared to $Lsh^{+/+}$ MEFs. In conclusion, specific protein coding genes and IAP retrotransposons gain H3K9me3 at day fourteen of OHT treatment, but not after OHT with UNC treatment. This indicates that the H3K9me3 mark observed is a result of H3K9me2 modification initially deposited by the G9a/GLP complex.

I also analysed nucleosome density to determine whether UNC treatment was affecting the changes in H3 occupancy observed upon LSH restoration. ChIP for H3 was carried out using the same cell lines after OHT and OHT with UNC treatment. We observed no clear differences between the OHT and OHT with UNC treated samples (Figure 5.9 B). In general, $Lsh^{off/off}$ MEFs followed the same trend under both treatments. We observed a delay in the recovery of H3 occupancy at promoters of *Gm9* and IAPs after fourteen days of OHT with UNC treatment compared to the OHT only treated samples. However, more biological replicates should be done to confirm this observation.

Taken together, these results demonstrate that H3K9me2 deposition by the G9a/GLP complex is required for LSH-dependent gene silencing of *Rhox* genes while H3K27me3 seems to be playing a transitory role in the silencing of IAP retrotransposons in the absence of G9a/GLP activity. However, more experiments should performed to determine the role of HDACs in LSH-dependent gene silencing since the inefficient VPA treatment combined with OHT did not permit us to study whether HDACs are strictly required before H3K9me2 deposition.

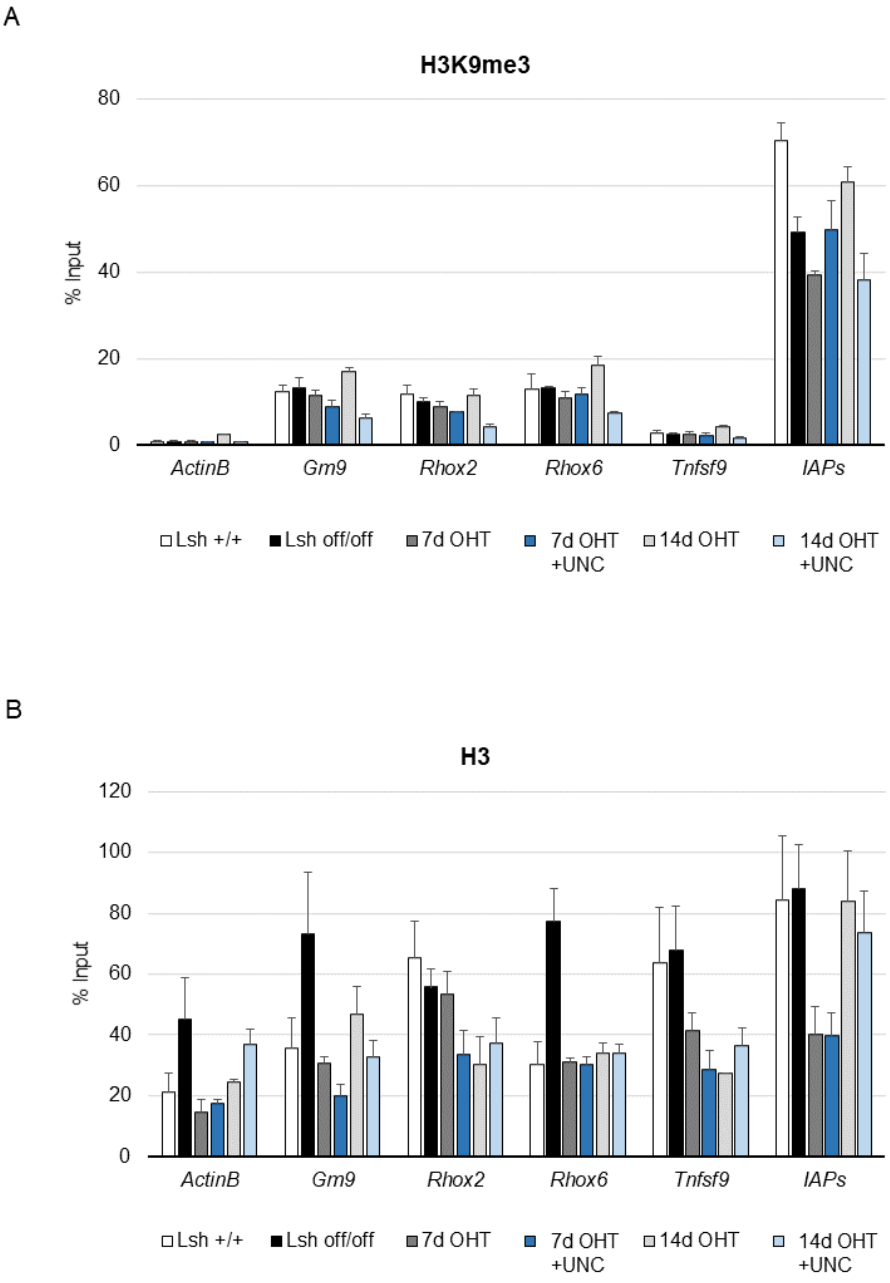


**Figure 5.9. Inhibition of G9a/GLP impairs the gain of H3K9me3 mark at LSH-dependent loci upon restoration of LSH but it does not affect H3 occupancy at promoters of specific genes and IAP retrotransposons.**

ChIP for H3K9me3 (A) and histone H3 (B) at specific protein coding genes and IAPs in *Lsh*$^{+/+}$, *Lsh*$^{off/off}$ and *Lsh*$^{off/off}$ MEFs after OHT and OHT with UNC treatment. *Actin* and *Tnfsf9* were used as controls since their transcriptional state does not change. The error bars represent standard error of the mean calculated from three technical replicates.

## 5.6 Discussion

To understand the role of LSH in epigenetic gene silencing it is important to consider the order of events taking place upon LSH restoration in the *Lsh*$^{off/off}$ MEFs. We have shown that after seven days of OHT treatment, when LSH protein levels are restored to wild-type, LSH localises at LSH-dependent loci leading to a mild decrease in H3 occupancy. This is accompanied by loss of H3Ac mark from these promoters at day seven of treatment, which happens at the same time the genes start to get silenced. At day fourteen of OHT treatment, LSH-dependent loci are silenced and we observe gain of H3K9me2 at their promoters. Surprisingly, all these events happen without gain of DNA methylation at the promoters of LSH-dependent genes and retrotransposons. This order of events occurring after expression of LSH in the knockout cells could indicate a primary role of LSH as a chromatin remodeling protein that also leads to removal of H3Ac mark, to establish a suitable substrate for G9a/GLP activity since these enzymes do not methylate acetylated histones. The subsequent methylation of H3K9 by G9a/GLP leads to a transcriptionally repressive chromatin state and DNA methylation could be involved in stabilizing this repressive state already established at LSH targeted regions.

LSH localises at all the studied loci after seven days of tamoxifen treatment, including control genes and LSH-dependent loci. This could indicate that after LSH restoration, LSH performs a genome-wide scanning but it only stays longer at specific places in the genome. The recently shown interaction of LSH with CDCA7 could target LSH to specific regions in the genome since CDCA7 is required for LSH recruitment to chromatin and contains a highly conserved zinc finger domain, which may facilitate their chromatin recruitment in a sequence specific manner (Jenness et al., 2018). The binding of LSH at these loci is accompanied by a mild decrease in H3 occupancy, which might indicate that LSH is acting as a chromatin remodeler either sliding nucleosomes or evicting the histone octamer. The sliding of nucleosomes by LSH could enable access to other proteins or protein complexes required for the generation of the transcriptionally repressive state, such as proteins that recruit HDACs to these target regions or histone demethylases that remove the H3K4me3 mark observed in the absence of LSH at LSH-dependent loci that gain transcription (Termanis et al., 2016). This nucleosome sliding could also facilitate binding of CDCA7 to specific sequences through its zinc finger domain and stabilise LSH recruitment. In addition to the nucleosome repositioning activity, LSH could be acting as a scaffolding protein to facilitate HDACs recruitment to these loci and removal of the acetylation mark (Myant & Stancheva, 2008), which is necessary for the subsequent deposition of H3K9me2 by the G9a/GLP complex. Another possibility for the chromatin remodeling activity of LSH could be the eviction of nucleosomes with active chromatin marks, which would allow the incorporation of non-acetylated histones to facilitate

the establishment of a repressive chromatin state by G9a/GLP activity. LSH has been recently shown to slide nucleosomes *in vitro* but free DNA was not generated during the course of the remodeling, indicating that the LSH-CDCA7 complex could not evict the nucleosome (Jenness et al., 2018). The optimisation of the HDACs inhibitor treatment will help clarify whether HDACs activity is essential for LSH-dependent gene silencing. This would rule out the hypothesis of LSH evicting nucleosomes with acetylated histones and reinforce the role of LSH as a chromatin remodeler that slides nucleosomes to facilitate the access of chromatin modifying enzymes that lead to LSH-dependent gene silencing.

We have also shown that DNA methylation is not necessary for the initial silencing of misregulated genes. In contrast to these findings, recent studies have shown that the role of LSH in gene silencing is accompanied by establishment of DNA methylation and this requires the ATPase activity of LSH (Ren et al., 2015; Termanis et al., 2016). Interestingly, these studies hypothesized a role for LSH as a chromatin remodeler in addition to its previously suggested role as a scaffolding protein (Myant & Stancheva, 2008). It has been previously shown in the lab that LSH cooperates with DNMT1, DNMT3B and the histone deacetylases HDAC1 and HDAC2 to silence transcription when LSH is targeted to the promoter of a reporter gene. This transcriptional repression by LSH and the interaction with HDACs are lost in DNMT1 and DNMT3B knockout cells. Histone deacetylation is required for silencing of the reporter gene, although the enzymatic activities of DNMTs are not required for this silencing mediated by LSH (Myant & Stancheva, 2008). Another publication also shows that p16 repression mediated by LSH requires HDACs activity, but it does not required DNMTs activity. Moreover, they also show that LSH associates with HDAC1 and HDAC2 (Zhou, Han, Li, & Tong, 2009). Accordingly, in our cellular system LSH could be sliding nucleosomes and also recruiting HDACs to promote the establishment of transcriptionally repressive state that might be further stabilized by DNA methylation at the targeted loci. It is not clear yet whether DNA methylation is rescued at some point after gene silencing is established in this conditionally reversible *Lsh* knockout cellular system. We would need to be able to select for the $Lsh^{on/on}$ cells in order to maintain the converted cells for a longer period in culture without the $Lsh^{off/off}$ cells outgrowing them. This would allow to study if DNA methylation is restored upon LSH expression and the specific time when it happens. It would also be interesting to know how long can this repressive state be maintained in the absence of DNA methylation.

After the removal of H3Ac upon LSH restoration, the next change in chromatin modifications observed is the deposition of H3K9me2 by G9a/GLP at LSH-dependent loci. The results from this chapter show that H3K9me2 deposition by G9a/GLP is required for LSH-dependent gene silencing of *Rhox* genes, but not of IAP retrotransposons. The histone modification H3K27me3 seems to play a transitory role in the silencing of IAP retrotransposons

in the absence of G9a/GLP activity. The gain of H3K27me3 after seven days of OHT treatment is not maintained during the fourteen days period of treatment although IAPs remained silenced at day fourteen of OHT treatment. Interestingly, PRC2-mediated histone methylation has been suggested to be necessary for PRC1 recruitment (Kuzmichev, Nishioka, Edrjument-Bromage, Tempst, & Reinberg, 2002). This transitory gain of H3K27me3 mark could be recruiting components of the PRC1 complex to stabilise this silencing through ubiquitylation of histone H2A (Boyer et al., 2006) and later promote DNA methylation to stably maintain silencing of IAPs. Reinforcing this hypothesis, LSH can associate with PRC1 components and influence PRC-mediated histone modifications such H2A ubiquitylation (Xi et al., 2007). Taken together, in the absence of G9a/GLP activity IAPs silencing mediated by PRC complexes could be facilitated by LSH-dependent nucleosome sliding and the interaction of LSH with different components of the complexes, promoting the association of PRC complexes with the target regions. As mentioned earlier, in contrast to the silencing of IAPs, *Rhox* genes require deposition of H3K9me2 by G9a/GLP for silencing upon LSH restoration. Previously it has been shown in the lab that cooperation between LSH and theG9a/GLP complex facilitates establishment of normal patterns of DNA methylation and gene silencing at specific promoters during differentiation of ES cells. Also, G9a/GLP recruitment is compromised when LSH is absent suggesting that LSH promotes the association of G9a/GLP with specific promoters although interaction between LSH and the G9a/GLP complex in coimmunoprecipitation experiments was not detected. This study suggested that LSH promotes G9a/GLP recruitment by bringing HDACs to remove acetylation that inhibits the activity of the G9a/GLP complex (Myant et al., 2011). The previous study carried out in the lab used differentiating ES cells with retinoic acid treatment (Myant et al., 2011). However, retinoic acid could be silencing these pluripotency genes in differentiating cells rather than LSH due to the stronger action of retinoic acid receptor. Here, we have shown that restoration of LSH expression in somatic cells leads to changes in chromatin structure and histone modifications that create a silence chromatin state without establishment of DNA methylation. We can hypothesize that removal of acetyl groups from H3 by LSH-associated histone deacetylases (Myant & Stancheva, 2008; Zhou et al., 2009) generates a suitable substrate for mono- and di-methylation of H3K9 and we observe subsequent gain of H3K9me2 at fourteen days of treatment after LSH localises at its targeted loci. In relation to this, mass spectrometry analysis carried out in the lab found LSH as an interacting partner of the G9a/GLP complex (T. Zhang et al., 2016). Therefore, binding of G9a/GLP to these loci could be facilitated by this suitable substrate generated and a possible interaction of LSH with the G9a/GLP complex.

In summary, we have demonstrated that LSH localises at LSH-dependent loci upon its restoration in our conditionally reversible *Lsh* knockout cellular system. We observe a mild

decrease in the occupancy of H3 at the promoters of these loci at the same moment that LSH localises there, reinforcing the previously shown role of LSH as a chromatin remodeler (Jenness et al., 2018). The activity of LSH as a remodeler sliding nucleosomes that could promote binding of other factors and the previously shown interaction of LSH with HDACs (Myant & Stancheva, 2008), facilitate the removal of acetyl groups from H3 tails observed at the same time point that LSH localises at these loci. Then, the suitable substrate created for G9a/GLP activity promotes deposition of H3K9me2 modification at the same time that misregulated genes are silenced. Taken together, our results suggest that the primary role of LSH is to promote changes in chromatin structure and modifications that lead to gene silencing and not DNA methylation, which most likely occurs as a consequence of transcriptional silencing.

# Chapter 6 - Conclusion and future outlook

Previous work has shown that specific genes misregulated in the *Lsh*$^{-/-}$ MEFs can be silenced and DNA methylation restored at their promoters in a cell autonomous manner when wild-type LSH is reintroduced into the knockout cells (Termanis et al., 2016). However, neither gene expression nor DNA methylation are rescued when the ATP mutant LSH is reintroduced. The silencing of these LSH-dependent genes is accompanied by loss of active chromatin marks, H3Ac and H3K4me3 (Termanis et al., 2016) from their promoter region. ATP hydrolysis by LSH is essential for its function in gene silencing and *de novo* DNA methylation (Ren et al., 2015; Termanis et al., 2016), which suggests that chromatin remodeling by LSH is necessary for these functions. From the previous study carried out in the laboratory we also know that the defects in gene expression and DNA methylation that occur in the absence of LSH in MEFs can be rescued autonomously (Termanis et al., 2016). Moreover, we have shown using the previous *Lsh* knockout system that an appropriate concentration of DNMT3B is necessary for the final rescue of gene repression and DNA methylation by wild-type LSH (Termanis et al., 2016). However, it is not clear the molecular mechanism of LSH-dependent gene silencing and *de novo* DNA methylation. We do not understand how LSH re-expression leads to the final repressive state, in which misregulated genes are silenced and *de novo* methylation established at their promoters. The principal aim of this work was to determine how LSH-dependent changes in chromatin structure and modifications facilitate epigenetic gene silencing and *de novo* DNA methylation. The use of the conditional knockout cellular system enabled us to express LSH in the *Lsh*$^{off/off}$ MEFs and study the order of events leading to the establishment of gene silencing and DNA methylation at LSH-dependent loci.

The characterization of the *Lsh*$^{off/off}$ MEFs is in agreement with previous studies demonstrating that LSH deficiency affects the expression and DNA methylation at promoters of specific single-copy protein coding genes (Myant et al., 2011; Tao et al., 2011; Xi et al., 2007) and repetitive elements (Dennis et al., 2001; Huang et al., 2004). Treatment with OHT for seven days is sufficient to convert the *Lsh*$^{off}$ allele to the *Lsh*$^{on}$ orientation in immortalised MEFs and the allele conversion restores LSH expression in the *Lsh*$^{on/on}$ MEFs to wild-type levels. Therefore, the conditionally reversible *Lsh* knockout cellular system functions as expected and allows us to study the order of events occurring immediately after LSH restoration in MEF cell lines in order to understand the molecular mechanism of LSH-dependent gene silencing. This was not possible to achieve in the earlier studies. However, since the non-converted cells outgrow the *Lsh*$^{on/on}$ MEFs, the changes in chromatin produced by LSH that lead to gene silencing and establishment of DNA methylation could only be

investigated in a narrow window of fourteen days. During the first seven days of OHT treatment, when LSH is restored, it localises at the promoters of LSH-dependent loci leading to a mild decrease in the occupancy of H3 and removal of acetyl groups from H3 tails. As these events occur at the same time that LSH binds to these loci, this reinforces the role of LSH as a chromatin remodeler (Jenness et al., 2018) and a scaffolding protein, since LSH has been previously shown to interact with HDACs (Myant & Stancheva, 2008; Zhou et al., 2009). Jenness et al have shown that CDCA7 is required to stimulate the ATPase-dependent nucleosome remodeling activity of LSH, which leads to nucleosome sliding. This nucleosome sliding could facilitate the recruitment of chromatin modifying enzymes such as histone demethylases to remove the H3K4me3 observed in *Lsh* knockout cells (Termanis et al., 2016). Then, the suitable substrate created after removal of H3Ac mark by HDACs interacting with LSH, facilitates the deposition of H3K9me2 by G9a/GLP histone methylases at the same time point when misregulated genes are completely silenced. This order of events suggests that LSH plays a role as a chromatin remodeler and also as a scaffolding protein leading to changes in chromatin structure and modifications that facilitate epigenetic gene silencing without rescue of DNA methylation in the initial period when LSH is restored in MEF cell lines (Figure 6.1). This supports earlier findings showing that transcriptional repression of the *Oct-4* gene during early embryogenesis is followed by an increase in H3K9me2 mediated by the G9a/GLP complex. The methylation of H3K9 promotes local heterochromatinization through binding of HP1 protein which is needed for subsequent *de novo* DNA methylation by DNMT3A/3B (Feldman et al., 2006). The G9a/GLP complex has also been shown to direct DNA methylation in mouse embryos. The silencing of germline genes mediated by G9a/GLP is incomplete without the establishment of DNA methylation by DNMT3B (Auclair et al., 2016). Moreover, H3K9me3 mediated by Suv39h has been shown to direct *de novo* DNA methylation via HP1 to pericentric repeats (Lehnertz et al., 2003). In relation to this, HP1 has been shown to bind both H3K9me2 and H3K9me3 marks via its chromodomains (Fischle et al., 2003). Therefore, a possible hypothesis for the establishment of DNA methylation in LSH-dependent loci might be the recruitment of DNMT3A/B through LSH, G9a or HP1 that binds to the already established methylation at lysine 9 of histone 3. However, *de novo* DNA methylation is not necessary for the establishment of the initial transcriptionally repressive chromatin state upon LSH restoration in somatic cells.

Further work is required to determine whether DNA methylation occurs in the reversible knockout cellular system to provide long term stability of the initial silent state created upon LSH restoration. The capacity to select for the $Lsh^{on/on}$ MEFs is needed in order to maintain these cells for a longer period without the non-converted cells outgrowing them after fourteen days in culture. This would allow to study the possible establishment of DNA methylation later

on time. It would be very interesting to know whether *de novo* DNA methylation follows the repressive state mediated by LSH restoration in these cells. In case there is rescue of DNA methylation following the repression, the next question to ask would be when it happens and how long can this repressive state be maintained in the absence of DNA methylation. A possible approach to select for the converted *Lsh*$^{on/on}$ MEFs would be to cross the mice containing the conditionally reversible *Lsh* cassette with mice containing a Cre-dependent fluorescent reporter. This would enable to distinguish the cells were recombination of the cassette occurred since the cells converting the GFPneo cassette to the *Lsh*$^{on}$ orientation will also express the fluorescent reporter, allowing us to select for the *Lsh*$^{on/on}$ MEFs by cell sorting. After selection of the *Lsh*$^{on/on}$ MEFs we could study whether DNA methylation is established following LSH-mediated repression at these loci. Once we have determine the time when DNA methylation occurs, we could perform ChIP experiments for G9a, HP1 and DNMT3B at different time points to clarify whether they play a role in the establishment of DNA methylation at these repressed loci.

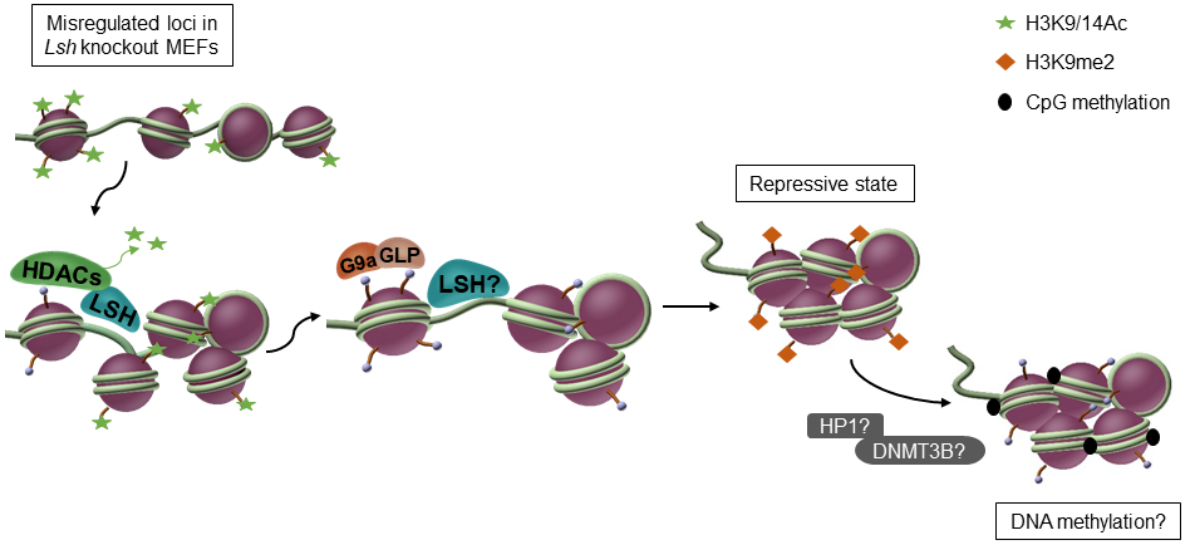


**Figure 6.1. Schematic diagram of the mechanism of LSH-dependent gene silencing in MEF cell lines.**

Model explaining the mechanism of LSH-dependent gene silencing upon LSH restoration in *Lsh*$^{off/off}$ MEFs. The model shows nucleosomes over the active misregulated genes in *Lsh*$^{off/off}$ MEFs. The promoters initially contain H3K9/K14Ac. When LSH is restored after OHT treatment of the *Lsh*$^{off/off}$ MEFs, it localises to these promoters potentially leading to sliding of nucleosomes, manifested by a decrease in H3 occupancy, and removal of H3Ac mark by HDACs interacting with LSH. Once deacetylated, H3K9 becomes a substrate for methylation by G9a leading to H3K9me2 deposition. However, it is not clear whether or not LSH interacts with G9a to facilitate its recruitment to these loci. H3K9me2 deposition is accompanied by silencing of misregulated genes creating a repressive chromatin state. Then, the protein HP1 could be recruited via H3K9me2 modification and promote *de novo* DNA methylation by DNMT3B enzyme, which could be recruited to the promoter through HP1, G9a or LSH. The establishment of appropriate patterns of DNA methylation following the gene repression could provide long term stability of the silent state.

Furthermore, the use of mass spectrometry to identify which proteins or protein complexes can be pulled down with LSH could provide more insight into the recruitment of G9a/GLP and DNMT3 enzymes by LSH to these LSH-dependent loci. This would improve our understanding of the role of LSH in gene silencing and establishment of *de novo* DNA methylation. To determine which LSH-dependent loci are able to re-establish normal patterns of expression upon LSH restoration, genome-wide analysis should be performed. To know which loci are silenced we should performed RNA-Seq followed by ChIP-Seq for LSH to attribute this silencing to the action of LSH. ChIP-Seq for LSH would also clarify whether it binds preferentially to specific regions of the genome, since it was shown that LSH is important for DNA methylation at the nuclear compartment that is in part defined by lamin B1 attachment sites (Yu et al., 2014). These approaches will help understand the dynamics of LSH-dependent transcriptional silencing in a genome-wide context. We have shown that LSH localises at the loci in study at day seven of OHT treatment when its protein expression levels are restored to wild-type. At the same time point there is a mild decrease in H3 occupancy at these promoters, but it is not clear how LSH remodels nucleosomes to enable changes in histone modifications. LSH could be repositioning nucleosomes or evicting the nucleosomes, which would also remove the acetylation mark and create a suitable substrate for G9a/GLP to methylate H3K9. The optimization of the HDACs inhibitor treatment would clarify whether HDACs activity is essential for this initial gene silencing or whether the acetylation mark is removed by nucleosome eviction. However, it is more probable that LSH acts sliding nucleosomes since it has been shown in *in vitro* assays that LSH can slide nucleosomes but not evict them (Jenness et al., 2018). The use of the NOMe-Seq assay at different time points after LSH restoration would clarify the dynamics of chromatin genome wide and its correlation with the *de novo* DNA methylation mediated by LSH. As mentioned earlier, recently it was demonstrated that LSH forms part of a bipartite nucleosome remodelling complex with CDCA7 (Jenness et al., 2018). The complex that LSH forms with CDCA7 possesses nucleosome repositioning activity although LSH alone cannot slide nucleosomes. Jenness et al have also shown that CDCA7 is essential for binding of LSH to chromatin in cells. It would be interesting to determine whether CDCA7 is necessary for recruitment of LSH to chromatin in our cellular system. CDCA7 could be knocked out in the $Lsh^{off/off}$ MEFs to study whether or not gene silencing occurs in the converted cells when LSH is restored in the absence of CDCA7. This could also help gain some insight into the requirement of ATP-dependent nucleosome remodeling by LSH in the initial repression of genes before acquisition of DNA methylation. It has been shown that the ATP binding by LSH is necessary for its function in *de novo* DNA methylation (Ren et al., 2015; Termanis et al., 2016), but it might not be required for the initial changes in chromatin modifications that lead to LSH-dependent gene silencing.

Several studies have shown that LSH plays a role in the establishment of *de novo* DNA methylation, but its role in the maintenance of DNA methylation has never been clarified. The sodium bisulfite DNA sequencing of IAPs from primary *Lsh*$^{off/off}$ MEFs and immortalised cells at different passage numbers did not show progressive loss of DNA methylation at higher passages. This preliminary analysis could indicate that LSH might not have a role in the maintenance of DNA methylation, although we have some *in vivo* data from mouse embryos (unpublished data) suggesting that LSH might support maintenance as well as *de novo* DNA methylation. The best experiment to analyse this would have been to convert the allele in the *Lsh*$^{on/on}$ MEFs to *Lsh*$^{off}$ orientation and study whether the normal patterns of DNA methylation characteristic of wild-type cells are lost in the absence of LSH. This would prove that LSH has a role in the maintenance of DNA methylation. I tried to perform this experiment using different cell lines but several issues arose with the cells and the conversion of the allele could not be optimised. Therefore the conversion of the allele to the *Lsh*$^{off}$ orientation needs to be optimised in order to study the possible role of LSH in maintenance of DNA methylation. The use of the conditional knockout system also allows the study in mouse embryos with the potential to reactivate LSH at different developmental stages which would provide a better characterization of the developmental timing of LSH-dependent events.

In conclusion, this work has demonstrated that changes in chromatin modifications leading to a repressive chromatin state can be established in somatic cells by the chromatin remodeler LSH to facilitate LSH-dependent gene silencing without acquisition of DNA methylation. This suggests that the primary role of LSH is to promote changes in chromatin structure and modifications that lead to gene silencing and not DNA methylation, which most likely occurs as a consequence of transcriptional silencing. This is an important aspect of LSH function since mutations in LSH and CDCA7 have recently been found to cause ICF syndrome (Thijssen et al., 2015), which is commonly associated with mutations in DNMT3B. The evidence from this study showing that the changes in chromatin required to silence misregulated genes can occur without gain of DNA methylation when LSH expression is restored in somatic cells, could lead to further research into new therapeutic options for patients with ICF syndrome.

# Appendix 1

*7592–7604* *Nucleic Acids Research, 2016, Vol. 44, No. 16* *Published online 13 May 2016*
*doi: 10.1093/nar/gkw424*

# The SNF2 family ATPase LSH promotes cell-autonomous *de novo* DNA methylation in somatic cells

**Ausma Termanis[1,†], Natalia Torrea[1,†], Jayne Culley[2], Alastair Kerr[1], Bernard Ramsahoye[2] and Irina Stancheva[1,*]**

[1]Wellcome Trust Centre for Cell Biology, University of Edinburgh, Michael Swann Building, Max Born Crescent, Edinburgh EH9 3BF, UK and [2]Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road South, Edinburgh EH4 2XR, UK

Received February 14, 2016; Revised May 02, 2016; Accepted May 04, 2016

## ABSTRACT

**Methylation of DNA at carbon 5 of cytosine is essential for mammalian development and implicated in transcriptional repression of genes and transposons. New patterns of DNA methylation characteristic of lineage-committed cells are established at the exit from pluripotency by *de novo* DNA methyltransferases enzymes, DNMT3A and DNMT3B, which are regulated by developmental signaling and require access to chromatin-organized DNA. Whether or not the capacity for *de novo* DNA methylation of developmentally regulated loci is preserved in differentiated somatic cells and can occur in the absence of exogenous signals is currently unknown. Here, we demonstrate that fibroblasts derived from chromatin remodeling ATPase LSH (HELLS)-null mouse embryos, which lack DNA methylation from centromeric repeats, transposons and a number of gene promoters, are capable of reestablishing DNA methylation and silencing of misregulated genes upon re-expression of LSH. We also show that the ability of LSH to bind ATP and the cellular concentration of DNMT3B are critical for cell-autonomous *de novo* DNA methylation in somatic cells. These data suggest the existence of cellular memory that persists in differentiated cells through many cell generations and changes in transcriptional state.**

## INTRODUCTION

Methylation of DNA at the fifth carbon of cytosine (5mC) is an abundant epigenetic modification in vertebrate genomes (1). In mammals, DNA methylation is established during development and contributes to regulation of genomic imprinting, tissue-specific gene expression, silencing of retrotransposons and X chromosome inactivation in females (2,3). The deposition of new methyl groups to cytosine occurs by the action of two homologous enzymes, the *de novo* DNA methyltransferases DNMT3A and DNMT3B, while the propagation of 5mC through DNA replication requires the activity of maintenance DNA methyltransferase DNMT1 (4). DNMTs are critical in early mammalian development when, following a nearly global erasure of 5mC during the cleavage stages of pre-implantation embryo, new patterns of 5mC are established post-implantation in the developing epiblast (E6.5) (3,5,6). Embryos lacking either DNMT1 or DNMT3B display severe 5mC deficiency and die at mid-gestation (E9.5–E11) (7,8). Several studies have identified DNMT3B as the main enzyme responsible for *de novo* DNA methylation during development (6,8–10). In *Dnmt3b−/−* embryos, the centromeric repeats, promoters of germ cell-specific genes and genes on the inactive X chromosome in female embryos remain hypomethylated.

The occurrence of new methylation at specific time of development suggests that the levels and the activity of DNMTs must be tightly controlled and coupled to developmental signaling. Several signal transduction pathways, in particular FGF and WNT, have been implicated in the exit from pluripotency, priming of embryonic cells for differentiation and regulation of DNA methylation. Thus simultaneous inhibition of mitogen-activated protein kinase (MAPK) and glycogen synthase kinase 3 (GSK3) pathways by specific inhibitors (2i) reinforces the naïve pluripotency of embryonic stem (ES) cells and this is accompanied by rapid downregulation of DNMT3B and loss of 5mC (11–13).

In addition to developmental signaling, the activity of DNMTs is also regulated at the level of chromatin. Unlike DNMT1 that methylates newly replicated hemimethylated DNA largely devoid of nucleosomes, the DNMT3 en-

*To whom correspondence should be addressed. Tel: +44 131 650 7029; Email: istancheva@ed.ac.uk
†These authors have contributed equally to this work as first authors.
Present address: Ausma Termanis, Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland.

zymes must function on DNA organized into chromatin. In comparison to naked DNA, stably positioned nucleosomes are a poor substrate for *de novo* DNA methylation *in vitro* and partly *in vivo* (14,15). Therefore the efficient *de novo* methylation of chromatin-organized DNA in cells and embryos requires either dynamic repositioning of nucleosomes or loosening of the contacts between the histones and DNA. In agreement with this, several ATP-dependent chromatin remodeling enzymes have been implicated in the regulation of 5mC levels and patterns, including the mammalian SNF2 family ATPases ATRX and LSH (16,17). A knockout of *Lsh* (*Hells*) gene in mice results in ~50% reduction of the global 5mC in the genome affecting repetitive sequences and large chromosomal domains associated with the nuclear lamina (16,18,19). Mapping of 5mC at promoters of protein-coding genes in wild-type and *Lsh*−/− mouse embryonic fibroblasts (MEFs) detected loss of 5mC from 20% of normally methylated promoters (19), many of which undergo lineage-specific silencing and *de novo* DNA methylation during early mouse development (10). Importantly, many of these genes are inappropriately expressed in the *Lsh*−/− MEFs (19). As DNMTs are present at normal levels in LSH-deficient cells (16) and LSH interacts directly with DNMT3B (20), these findings suggest that ATP-dependent chromatin remodeling is critical during development to open up chromatin for developmentally programmed DNA methylation by *de novo* enzymes.

If the programmed *de novo* DNA methylation were tightly regulated by signaling pathways in the developing embryo, one would predict that the loss of 5mC would be irreversible in somatic cells taken out of their normal developmental context. In order to investigate whether this is the case, we restored the expression of LSH in spontaneously immortalized hypomethylated *Lsh*−/− MEFs grown in culture for many cell generations. Contrary to our expectations, we found that 5mC at repetitive and unique sequences as well as gene silencing of developmentally regulated loci could be substantially reestablished when a wild-type LSH protein was introduced into the *Lsh*−/− MEFs. We also found that the reversal of 5mC levels and patterns in the *Lsh*−/− MEFs required the catalytic activity of LSH ATPase and appropriate cellular concentration of DNMT3B. Taken together, these experiments demonstrate that the capacity for LSH-regulated *de novo* DNA methylation of repetitive sequences and transcriptionally active developmentally regulated promoters is preserved in somatic cells. These experiments also suggest the existence of epigenetic cellular memory, which persists through changes in transcriptional state and permits silencing of inappropriately expressed genes and repetitive sequences in the absence of developmental signaling.

## MATERIALS AND METHODS

### Cell culture

All mouse embryonic fibroblast cell lines were grown in DMEM (Sigma) supplemented with 10% fetal bovine serum (Thermo Fisher), 100 units/ml penicillin, 1 mg/ml of streptomycin and 2 mM L-glutamine.

### Establishment of stable cell lines

pMSCV-LSH and pMSCV-K237Q plasmids were generated by cloning into pMSCV vectors carrying either puromycin or hygromycin resistance marker of mouse synthetic LSH ORF (wild-type and mutant; Life Technologies/GeneART) designed to contain an in-frame C-terminal 3xFLAG tag followed by the 3′UTR and polyadenylation sequence of the LSH mRNA. pMSCV, pMSCV-LSH and pMSCV-K237Q plasmids were packaged into lentiviral particles in Phoenix A cells. Culture supernatants were harvested 48 h post-transfection and the lentiviral titres determined by infection of NIH-3T3 mouse fibroblasts. $3 \times 10^5$ *Lsh*−/− MEFs were infected with lentiviral particles at MOI = 1 in the presence of 4 μg/ml Polybrene. The cells were transferred to larger dishes 48 h post-infection and selected with 2.5 μg/ml puromycin for 7–10 days. Individual colonies were isolated and expanded into cell lines. DNMT3B was knocked down by infecting the *Lsh*−/− MEFs with packaged lentiviral particles (Thermo Scientific) carrying pGIPZ plasmids with either non-silencing shRNAmir (V3LMM_420607) or shRNAmir sequences targeting murine DNMT3B mRNA (V2LMM_257108; V2LMM_257108; V2LMM_53727) at MOI = 3 in the presence of 4 μg/ml Polybrene. The cells were selected with 2.5 μg/ml puromycin for 7–10 days and subsequently infected with lentiviral particles carrying pMSCV-LSH-hygro at MOI = 1. The cells were selected with 400 μg/ml hygromycin and single colonies expanded into cell lines.

### Western blots

Nuclear extracts were prepared as described (19) and 60 μg of each extract were resolved in 7% SDS-polyacrylamide gel. The gels were transferred to nitrocellulose membrane (BioRad), blocked with 4% skimmed milk in 1 × TBS; 0.1% Tween and incubated with primary antibodies in blocking solution overnight. The membranes were washed extensively with PBS; 0.1% Tween, re-blocked for 30 min and incubated with secondary anti-mouse IR-800 and anti-rabbit IR-670LT antibodies (LiCOR Biosciences) for 1 h. After washing with PBS; 0.1% Tween, the membranes were imaged on Odyssey scanner (LiCOR Biosciences) and quantified when required with Image Studio Software (LiCOR Biosciences). The primary antibodies used for Western blots were: anti-LSH (Santa Cruz; sc-46665), anti-Flag M2 (Sigma), anti-MRE11 (Calbiochem; PC388), anti-HDAC1 (Santa Cruz; sc-7872) and anti-DNMT3B (Thermo Fisher; AFPA1884).

### Southern blots

Genomic DNA was purified from MEF cell lines using standard protocols and 10 μg of each DNA were digested with either MspI or its methylation-sensitive isoschizomer HpaII for 3 h at 37°C. DNA was resolved on 1% Agarose gels run in 1 × TAE buffer, the gels were transferred to nitrocellulose membrane (Pall B; VWR) with 0.4 M NaOH and hybridized with minor-satellite probe labeled with $\alpha$-$P^{32}$-dCTP (Perkin Elmer) by a random-priming kit (Thermo Scientific) in buffer containing 1 mM EDTA, 0.5

M NaHPO4 and 7% SDS at 65°C overnight. The blots were subsequently washed with 3× SSC; 0.1% SDS at 65°C and imaged on Typhoon phosphorimager (GE Healthcare).

### Bisulfite DNA sequencing

For bisulfite DNA sequencing, genomic DNA was processed with EpiTect Bisulfite conversion kit (QIAGEN) and used as a template for PCRs with specific primer pairs, which are listed in Supplementary Table S1. PCR products were cloned into pJet vector (Thermo Scientific) and sequenced using a reverse pJet primer and BigDye sequencing mix (Applied Biosystems). Sequences were analyzed by BiQ Analyzer software (21).

### Analyses of DNA methylation by reverse-phase HPLC

DNA was extracted from ESCs following standard protocols and residual RNA was removed by enzymatic hydrolysis (6-hour incubation with RNase A and RNase T1) followed by DNA precipitation in 3 volumes of ethanol. This was repeated once. 1–5 μg of purified DNA was digested with DNase1 (New England Biolabs) for 12 h in the recommended buffer. Following this, 2 volumes of 30 mM sodium acetate pH 5.2 was added and the DNA was further digested to nucleotide 5′ monophosphates with Nuclease P1 (Sigma) in the presence of 1 mM zinc sulphate (7-h incubation). The quantitation of 5-methylcytosine in genomic DNA was performed in triplicate by isocratic high performance reverse phase liquid chromatography as previously described (22), with the following alterations. A Dionex UM 3000 HPLC system was used complete with a column chiller, C18 column (250 mm × 4.6 mm 5 um APEX ODS, #4M25310, Grace Discovery Sciences), and column guard (Phomenex, #AJ0-7596). The mobile phase was 50 mM ammomium phosphate (monobasic) pH4.1. The column was chilled to 8°C to improve peak separation. Deoxyribonucleotides (dNMPs) were detected at their extinction maxima using a Dionex 3000 multiple wavelength detector: dCMP, 276 nm; 5mdCMP, 282 nm. Quantifications were calculated from the area under each peak using the respective extinction coefficients (dCMP, $8.86 \times 10^3$; 5mdCMP $9.0 \times 10^3$).

### Quantitative RT-PCR

Total RNA was purified from cell lines using TRIzol reagent (Life Technologies) according to manufacturer's instructions and cDNA synthesis performed with Superscript II (Life Technologies). Quantitative reverse transcription PCRs were carried out in three biological replicates with six technical replicates for each with SYBR-Green Mastermix (Roche) on Roche LC480 instrument. Fold changes relative to *GAPDH* were calculated using the Pfaffl method (23). Primer sequences are listed in Supplementary Table S1.

### Chromatin immunoprecipitation

Chromatin immunoprecipitation was performed essentially as described in (19). Briefly, the cells were crosslinked with 1% formaldehyde in DMEM supplemented with 10× crosslinking buffer (500 mM HEPES, pH7.9; 1.5 M NaCl; 10 mM EDTA; 5 mM EGTA), to 1× final concentration. After neutralising the formaldehyde with 2.5 M glycine (final concentration 125 mM), the cells were spun at 1200 rpm at 4°C and washed with PBS supplemented with PMSF. The crosslinked cells were resuspended and incubated in buffer L1 (50 mM HEPES, pH 7.9; 140 mM NaCl; 1 mM EDTA; 10% glycerol; 0.5% NP-40; 0.25% Triton X-100) for 5 min, spun and resuspended in buffer L2 (10 mM Tris pH 8; 200 mM NaCl; 1 mM EDTA; 0.5 mM EGTA). After incubation for 5 min, the cells were spun again, resuspended in TE with 0.3% SDS at density $4 \times 10^6$ cells/ml and sheared with BioRuptor sonicator (Diagenode) to fragments of average size 250–300 bp. 10 μg of chromatin were used per IP with 4 μg of either rabbit anti-H3K4me3 antibody (Active Motif) or control rabbit IgG. IP washes were performed as described elsewhere (19). After descrosslinking, immunoprecipitated DNA was purified with PureLink kit (Thermo Fisher) and 1/50 of input and ChIP DNA were used in each 20 μl quantitative PCR reaction. ChIP experiments were performed in two biological replicates with six technical replicates for each. Primers are listed in Supplementary Table S1.

### MeDIP sequencing and data analyses

MeDIP was performed essentially as described (24) with anti-5mC antibody (Eurogentec) on genomic DNA sonicated to average fragment size 300 bp. Ligation of Illumina adaptor sequences was performed prior to MeDIP with NEB-next DNA library Mastermix Set and Singleplex Illumina adaptor oligos (NEB) according to manufacturer instructions. After MeDIP, the DNA was amplified with 14 cycles of PCR and size selected on agarose gel for fragments 250–350 bp. The resulting libraries were Single-end sequenced on Illumina HiSeq2000 instrument generating 120–140 million reads per sample 99% of which could be mapped to the mouse genome. The MeDIP-seq reads were mapped with Burrows Wheeler Algorithm mem version 0.7.5a to the mm10 genome assembly. Sam files were processed into bam files using samtools version 0.1.19. Count tables for each of the studied regions were created using the coverage program in version 2.17.0 of the bedtools suite. Count tables were first normalized to raw reads per kilobase of feature, and then normalized to regions that were previously shown not to change DNA methylation between wild-type and *Lsh*−/− MEFs (19). A linear regression line of the log2 read counts of these non-changing regions was calculated, and the whole dataset was transformed such so that the line would have an intercept of 0 and a gradient of 1. A script written in R (normalise2file.r) can be downloaded from https://github.com/AlastairKerr/Termanis2015. Data was plotted using the ggplot2 package in R. Methylation recovery (MR) was calculated as MR = [100 x ($\log_2 D$ – $\log_2 S$)]/$D$ where $D$ = $\log_2$ read count WT - $\log_2$ read count MSCV and $S$ denotes the $\log_2$ read count for each of the cell lines expressing either LSH or K237Q. The MeDIP-seq data can be accessed at NCBI GEO database using accession number [GEO: GSE64384].

## RESULTS

### *Lsh*−/− MEFs expressing wild-type LSH display an increase in 5mC and methylate repetitive sequences

To investigate if the reduced levels of 5mC and the misexpression of normally silenced loci (genes and retrotransposons) are reversible in the *Lsh*−/− MEFs and if such reversal requires the catalytic activity of LSH, we introduced by lentiviral transduction a triple FLAG (3xFLAG)-tagged full length LSH, either wild-type or ATP binding-deficient mutant (25), carrying lysine 237 mutated to glutamine (K237Q), into the *Lsh*−/− MEFs. As an additional control, we also infected the LSH-null cells with viral particles carrying an empty vector (pMSCV). As the lentiviral vector carries puromycin resistance marker, we subjected the transduced cells to selection and subsequently isolated stable clonal cell lines (Figure 1A). For convenience, we will refer to these cell lines as *Lsh*−/− LSH, *Lsh*−/− K237Q and *Lsh*−/− MSCV. Quantitative Western blots (Figure 1B and C) and immunostaining with anti-FLAG antibodies (Figure 1D) revealed that most cell lines, except *Lsh*−/− MSCV, stably expressed either the wild-type or mutant LSH, although at different levels, and displayed a relatively homogeneous distribution of the 3xFLAG-LSH throughout the nucleus. We selected four cell lines, two expressing the wild-type LSH (Figure 1B and C, labeled 1 and 3) and two expressing the K237Q mutant (Figure 1B and C, labeled 3 and 4), for further analyses.

To examine if the expression of either LSH or K237Q in the *Lsh*−/− MEFs led to detectable changes in DNA methylation, we analyzed by High Performance Liquid Chromatography (HPLC) the total amount of 5mC in DNA purified from the clonal cell lines and compared the values with those of the controls, *Lsh*−/− MSCV and wild-type MEFs. The HPLC analyses detected a significant increase (~25%) of 5mC in both *Lsh*−/− LSH cells lines compared to *Lsh*−/− MSCV, but not in the *Lsh*−/− K237Q cell lines (Figure 2A). These data suggest that the wild-type LSH, but not the catalytically inactive K237Q mutant, promotes *de novo* DNA methylation in the *Lsh*−/− MEFs although both proteins and DNMT3B displayed comparable association with chromatin (Supplementary Figure S1).

DNA methylation of most repetitive sequences, including minor (centromeric) and major (pericentric) satellite repeats, IAP retrotransposons and LINE elements, is LSH-dependent (26) (Figure 2B and C, compare WT with MSCV) and some of these sequences are transcribed in the *Lsh*−/−, but not in wild-type MEFs (27,28). To investigate further the patterns of DNA methylation in the *Lsh*−/− MEFs with restored LSH expression, we analyzed DNA methylation at specific repetitive sequences. Southern blots on DNA digested with methylation sensitive HpaII restriction enzyme hybridized with minor satellite probe detected the characteristic ladder of hypomethylated satellite repeats in *Lsh*−/− MSCV MEFs and in three independent *Lsh*−/− K237Q cell lines (Figure 2B). This laddering was largely absent in HpaII-digested DNA derived from *Lsh*−/− LSH cell lines (Figure 2B). Interestingly, there was no significant correlation between the efficiency of methylation at satellite sequences and the expression levels of wild-type LSH in the *Lsh*−/− cell lines (Compare Figure 1B and C with Figure 2B). Furthermore, bisulfite DNA sequencing of IAP retrotransposons detected a significant 20–40% increase in DNA methylation in the *Lsh*−/− LSH cell lines, while methylation of IAPs in *Lsh*−/− K237Q cells was comparable to that of the MSCV control (Figure 2C). Collectively, these analyses demonstrate that the *Lsh*−/− somatic cells are capable of efficient *de novo* DNA methylation of repetitive sequences upon expression of wild-type LSH, but not of the catalytically-inactive K237Q mutant.

### *De novo* gene silencing and DNA methylation in the *Lsh*−/− MEFs require LSH, but can occur in the absence of differentiation signals

As mentioned earlier, about 20% of all promoters normally methylated in wild-type MEFs lack DNA methylation in the *Lsh*−/− MEFs (19). Among these are the promoters of reproductive homeobox (*Rhox*) genes and pluripotency-associated genes (*Dppa2* and *Dppa4*), which undergo transcriptional silencing and programmed DNMT3A/3B-dependent *de novo* DNA methylation during early development in cells that contribute to the embryo proper (10,29). Normally, the *Rhox* genes are expressed and their promoters are not methylated in extraembryonic tissues and in germ cells, but are stably silenced and methylated in the embryo and all somatic cells (10,30). The vast majority of *Rhox* genes are hypomethylated and expressed in *Lsh*−/− MEFs, but not in their wild-type counterparts (19).

Given that the 5mC at the *Rhox* loci, including promoters, is established early in development following the specification of embryonic and extraembryonic lineages (10), we asked whether or not the *Lsh*−/− MEFs could methylate and silence these genes in the absence of exogenous signals. Surprisingly, bisulfite DNA sequencing detected an almost complete methylation of *Rhox2a, Rhox6* and *Rhox9* promoters in *Lsh*−/− LSH cell lines, while DNA methylation was absent in *Lsh*−/− K237Q and *Lsh*−/− MSCV cells (Figure 3). *De novo* DNA methylation of gene promoters in the *Lsh*−/− LSH cell lines was also accompanied by robust silencing of *Rhox* as well as other LSH-dependent loci, such as *Gm9* and IAP retrotransposons (Figure 4A). Transcriptional silencing and *de novo* DNA methylation of gene promoters in the *Lsh*−/− LSH cells were also accompanied by loss of histone H3 lysine 4 trimethylation (H3K4me3) (Figure 4B), a histone modification that marks active promoters and CpG islands in mammalian cells (31,32). None of the examined genes were silenced and often showed a mild increase in expression and H3K4me3 in the *Lsh*−/− K237Q MEFs (Figure 4A and B). In contrast to the *Rhox* loci, we did not observe silencing of *HoxC6, Peg12* and pluripotency genes *Dppa2* and *Dppa4* (Figure 4A). The promoters of *Dppa2* and *Dppa4* remained unmethylated in *Lsh*−/− LSH cells (Supplementary Figure S2A and B) suggesting that additional signals might be required for silencing and *de novo* DNA methylation of these loci. However, treatment of *Lsh*−/− LSH and *Lsh*−/− K237Q MEFs with retinoic acid induced silencing of *Dppa2* and *Dppa4* and methylation of promoter sequences, but this repression could not be stably maintained in *Lsh*−/− K237Q MEFs (Supplementary Figure S2C and D).

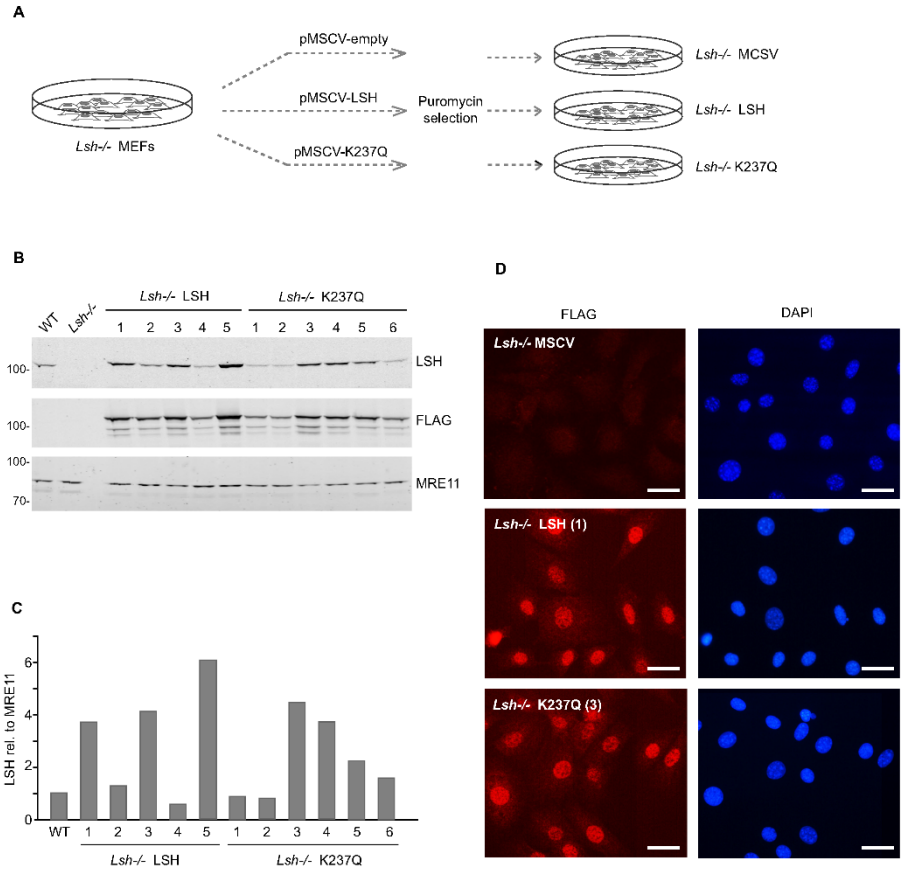


**Figure 1.** Generation of stable *Lsh*−/− MEF cell lines expressing either wild-type or catalytically inactive LSH. (**A**) Schematic diagram showing the generation of stable cell lines from the *Lsh*−/− MEFs by lentiviral transduction of packaged pMSCV vectors. (**B**) Expression of the wild-type LSH and ATP binding-deficient mutant form, LSH K237Q (K237Q), in clonal cell lines. Quantitative Western blots were probed with the indicated antibodies. MRE11 was used as a loading control. (**C**) Quantification of the wild-type and mutant 3xFLAG-tagged LSH protein levels in clonal cell lines. (**D**) Immunofluorescence staining of stable cell lines with anti-FLAG antibody. The nuclei were counterstained with DAPI. The scale bar corresponds to 20 μm.

Taken together, our analyses demonstrate that, in addition to repetitive sequences, the *Lsh*−/− LSH MEFs are capable of efficient silencing and *de novo* methylation of active, H3K4me3-marked developmentally regulated promoters and that these events require ATP hydrolysis by LSH and can take place in the absence of exogenous signals.

### The catalytic activity of LSH supports *de novo* DNA methylation by both DNMT3 enzymes

As the expression of wild-type LSH in the *Lsh*−/− MEFs restored the 5mC levels from 52% to ∼ 75% of those measured in wild-type MEFs (Figure 2A), we next asked whether or not the *Lsh*−/− LSH MEF cell lines displayed heterogeneity of DNA methylation patterns and if DNA methylation was restored uniformly across the genome. To address these questions, we carried out methylated DNA immunoprecipitation (24) followed by high-throughput sequencing (MeDIP-seq) on DNA purified from wild-type MEFs, *Lsh*−/− MSCV, *Lsh*−/− LSH (lines 1 and 3) and *Lsh*−/− K237Q (lines 3 and 4). The obtained reads were aligned to the genome and analyzed further. A visual inspection of DNA methylation patterns at specific loci, for example the *Rhox* cluster of homeobox genes (Figure 5A), and pairwise comparisons between the samples (Figure 5B) revealed that the DNA methylation patterns of both *Lsh*−/− LSH cell lines were very similar to each other and resembled those of the wild-type MEFs. As expected, the

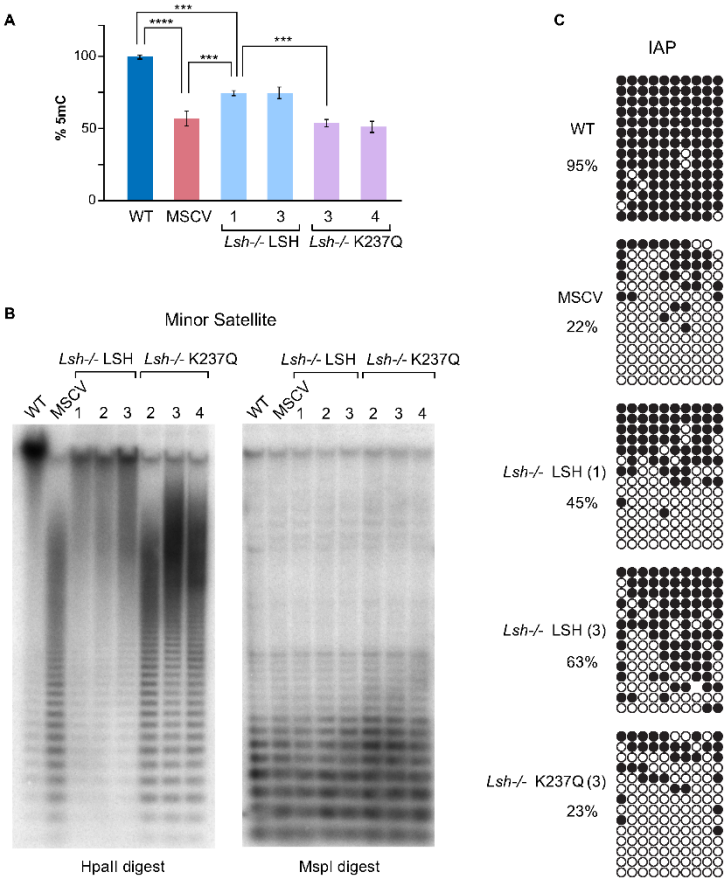


**Figure 2.** Wild-type LSH restores DNA methylation globally and at repetitive sequences in *Lsh*−/− MEFs. (**A**) Quantification of 5mC by HPLC in indicated cell lines. To simplify the comparison, the amount of 5mC in wild-type MEFs (3.3% of total cytosine) is represented as 100%. Error bars represent standard deviation, $n = 3$, $P$ values ($t$-test) **** is $P < 10^{-4}$ and *** is $P < 10^{-3}$. (**B**) Southern blots of either HpaII (methylation sensitive) or MspI (methylation insensitive) restriction enzymes-digested genomic DNA from indicated cell lines hybridized with $^{32}$P-labeled centromeric mouse minor satellite probe. (**C**) Bisulfite DNA sequencing of IAP retrotransposons in indicated cell lines. Each row represents a single DNA strand, and each circle is a CpG dinucleotide. Methylated CpGs are shown in black.

patterns of 5mC of the *Lsh*−/− K237Q cell lines showed the highest correlation with each other and the *Lsh*−/− MSCV cells. This suggests that the expression of LSH in the *Lsh*−/− MEFs results in a relatively uniform, although incomplete (Figures 2C and 3), *de novo* methylation of LSH-dependent loci.

In ES cells, DNMT3B is recruited to chromatin via binding to H3K36me3, a modification established co-transcriptionally in gene bodies, while DNMT3A displays a more broad distribution across the genome, which positively correlates with CpG density (15). In addition, both DNMT3A and DNMT3B, are capable of interacting with the unmodified histone H3 tails via a conserved ADD domain (33–35). Similar to ES cells, the calculated mean methylation of gene bodies in wild-type MEFs was higher than the median DNA methylation in the intergenic regions (Supplementary Figure S2A, WT) and methylation of both types of sequences was reduced proportionally in the *Lsh*−/− MEFs (Supplementary Figure S3A, MSCV).

To assess further if LSH supports *de novo* methylation by both DNMT3 enzymes, we calculated the mean DNA methylation recovery in gene bodies and intergenic regions for the *Lsh*−/− LSH and *Lsh*−/− K237Q cells (see Methods). The calculated recovery of DNA methylation was 40–

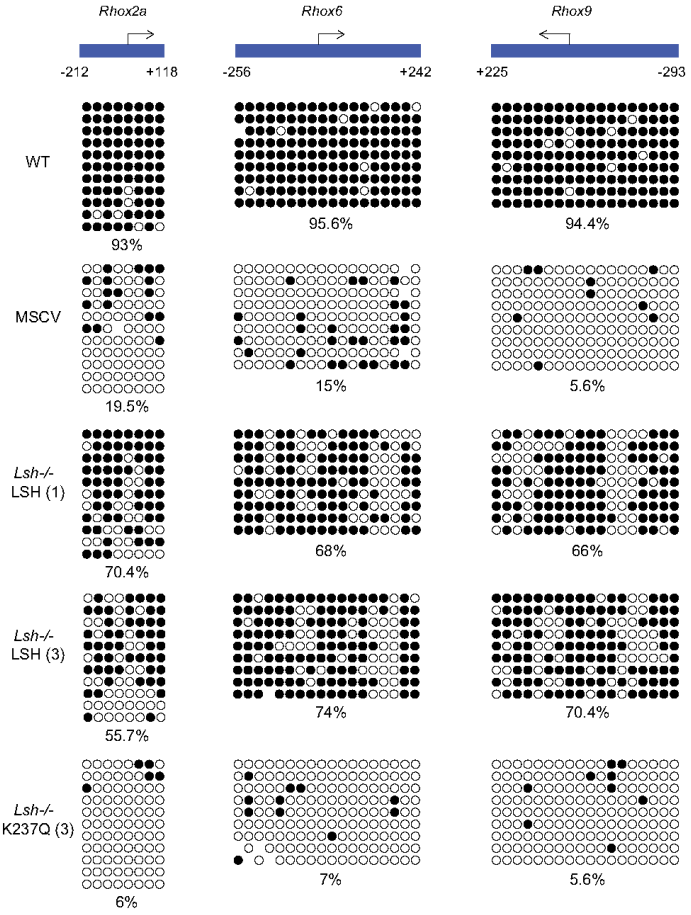


**Figure 3.** *De novo* DNA methylation of developmentally-regulated gene promoters in *Lsh*−/− MEFs expressing wild-type LSH. Bisulfite DNA sequencing of reproductive homeobox (*Rhox*) gene promoters in wild-type MEFs, *Lsh*−/− MEFs carrying an empty pMSCV vector and *Lsh*−/− MEFs expressing either wild-type or mutant LSH. The numbers below the promoter graphs indicate bp upstream (−) or downstream (+) from the transcription start site. Each row of circles is a single DNA strand and methylated CpGs are shown in black.

80% for both types of regions with higher median value for gene bodies than intergenic regions (Figure 5D). Interestingly, we also detected a weak, but consistent recovery of DNA methylation in the *Lsh*−/− K237Q cell lines (Figure 5D and Supplementary Figure S3). In contrast to wild-type LSH, the median recovery values for intergenic regions were higher than those for gene bodies in LSH K237Q expressing cells (Figure 5D). This indicates that although both DNMT3 enzymes require the catalytic activity of LSH for efficient *de novo* DNA methylation, the chromatin remodeling by LSH plays a more significant role in DNMT3B- rather than in DNMT3A-dependent methylation.

Consistent with the bisulfite sequencing data for IAPs and *Rhox* promoters (Figures 2 and 3), DNA methylation was also restored with 3–4 fold higher efficiency at all classes of repetitive sequences and gene promoters in the *Lsh*−/− LSH cell lines in comparison with *Lsh*−/− K237Q cells (Figure 5C, E and Supplementary Figure S3B). Overall, 96% of all promoters that were hypomethylated in the *Lsh*−/− MEFs substantially recovered 5mC in the *Lsh*−/− LSH cell lines (Supplementary Figure S3B and C). Taken together, these data support our global and locus-specific analyses of DNA methylation (Figures 1 and 2) and demonstrate that upon LSH expression in the *Lsh*−/− MEFs *de*

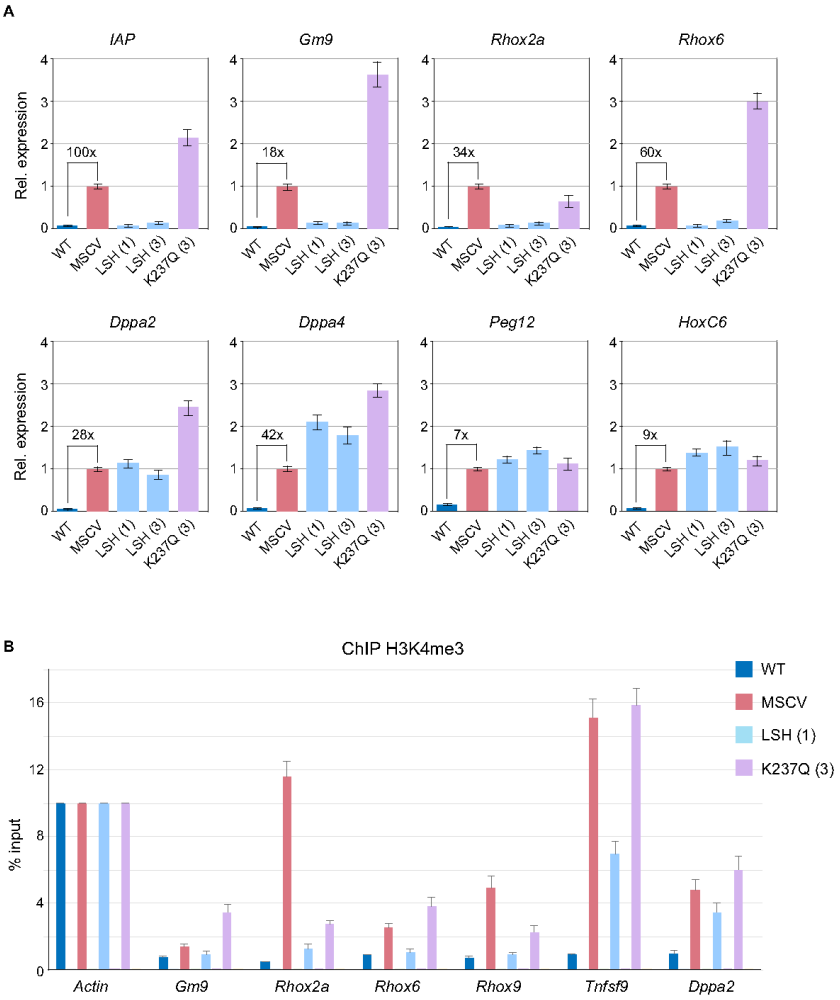


**Figure 4.** The *Lsh*−/− MEFs expressing wild-type LSH efficiently silence IAP retrotransposons and developmentally regulated genes. (**A**) Expression of IAP retrotransposons and specific genes in wild-type MEFs, Lsh−/− MEFs carrying an empty pMSCV vector and *Lsh*−/− MEFs expressing either wild-type or mutant LSH as assessed by quantitative RT-PCR. The bar graphs show fold change in expression relative to *Lsh*−/− MSCV MEFs. The housekeeping *Gapdh* gene was used for normalisation. The fold upregulation of gene expression detected by qRT-PCR in *Lsh*−/− MSCV relative to wild-type MEFs is shown numerically above the brackets. The error bars represent standard deviation. (**B**) Silencing of gene expression by wild-type LSH is accompanied by loss of H3K4me3 from gene promoters. *Tnfsf9* was randomly chosen from the list of promoters that aquire DNA methylation in the *Lsh*−/− LSH cells. The error bars represent standard deviation. A rabbit IgG was used as additional control for the anti-H3K4me3 antibody (not shown).

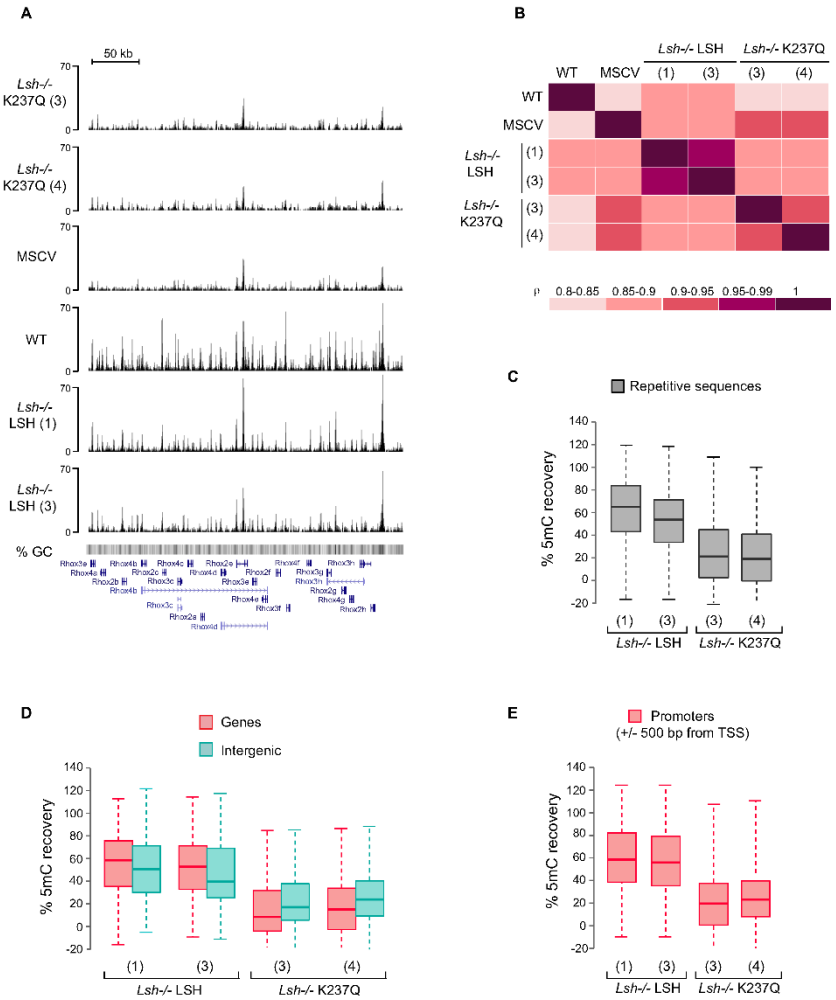


**Figure 5.** Global patterns of DNA methylation in the *Lsh−/−* MEFs expressing either wild-type or mutant LSH. (**A**) Profile of unique MeDIP-seq reads across a 330 kb region within the *Rhox* gene cluster on the X chromosome. (**B**) Spearman correlation coefficient matrix displays pairwise comparison of DNA methylation profiles between the analyzed cell lines. Boxplots display DNA methylation recovery (**C**) at repetitive sequences; (**D**) within genes and intergenic regions; (**E**) at gene promoters in the indicated cell lines. Only promoters methylated in wild-type MEFs, but not in the *Lsh−/−* MEFs were used in the graph shown in **E**. Each box in **C**, **D** and **E** represents the interquartile range (IQR) and the lowest and highest values within 1.5x IQR are shown as whiskers.

*novo* DNMTs are able to access genic, intergenic and promoter sequences.

### DNMT3B is necessary for *de novo* DNA methylation and silencing of LSH-dependent loci

DNA methyltransferase DNMT3B is implicated in the establishment of DNA methylation in post-implantation embryos (6,9) as well as in *de novo* methylation of the *Rhox* loci (10) and centromeric repeats (8). As LSH directly in-

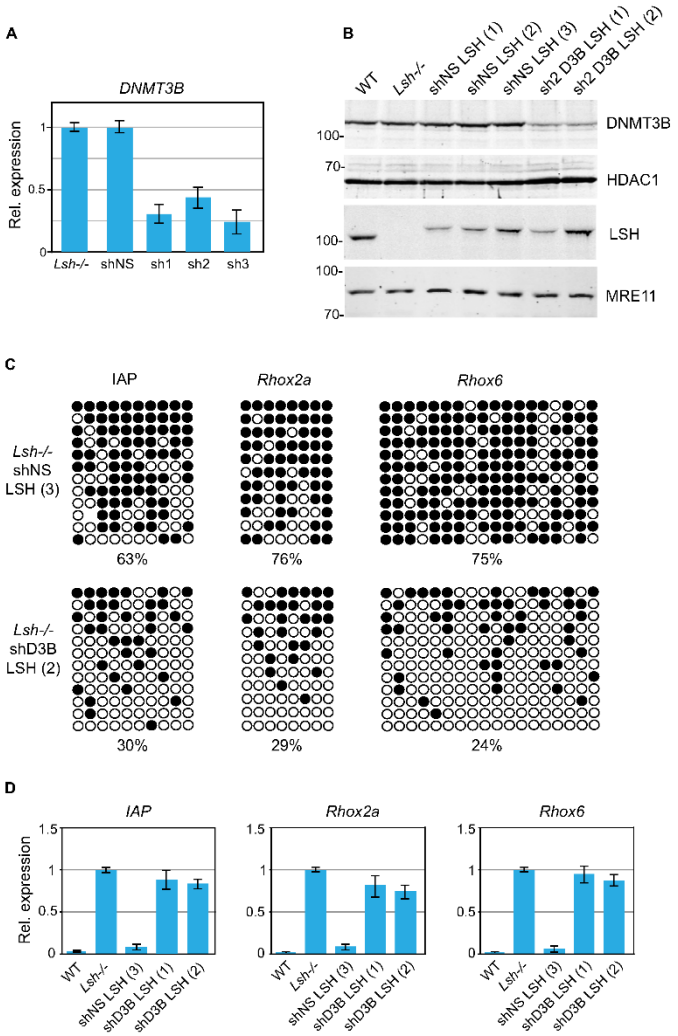


**Figure 6.** DNMT3B is required for *de novo* DNA methylation of IAPs and gene promoters upon LSH expression in *Lsh*−/− MEFs. (**A**) Expression of DNMT3B in *Lsh*−/− MEFs expressing either a non-silencing shRNA (shNS) or shRNAs (sh1, sh2 and sh3) targeting DNMT3B as assessed by quantitative RT-PCR. The error bars denote standard deviation. (**B**) Western blots showing the presence of 3xFLAG-tagged wild-type LSH in *Lsh*−/− clonal MEF lines expressing either non-silencing shRNA (shNS) or shRNA targeting DNMT3B (sh2 D3B). HDAC1 and MRE11 were used as loading controls. (**C**) Bisulfite DNA sequencing of IAPs and the promoters of *Rhox2a* and *Rhox6* genes in indicated cell lines. Methylated CpGs are shown in black. (**D**) Expression relative to *Gapdh* of *IAPs*, *Rhox2a* and *Rhox6* in indicated cell lines. The expression detected in *Lsh*−/− MEFs was designated as 1. The error bars denote standard deviation.

teracts with DNMT3B and indirectly, via DNMT3B, with DNMT1 and histone deacetylases HDAC1 and HDAC2 (20), we asked if DNMT3B is responsible for methylation of repetitive sequences and unique loci in *Lsh*−/− LSH MEFs. To address this, we first stably knocked down DNMT3B in *Lsh*−/− MEFs by lentiviral vectors expressing small hairpin (sh) RNAs (Figure 6A, sh1-3). In parallel, we established a control cell line carrying a non-silencing shRNA (Figure 6A, shNS). Although all three shRNA targeting DNMT3B reduced the levels of DNMT3B mRNA and protein by 60–75% (Figure 6A and B), cells carrying the most efficient shRNAs (sh1 and sh3) displayed growth inhibition and were not used further. We next introduced the wild-type LSH into the *Lsh*−/− sh2-D3B and into the control shNS cells and subsequently established stable clonal sh2D3B-LSH and shNS-LSH cell lines (Figure 6B).

Analyses of IAPs as well as *Rhox2* and *Rhox6* promoters by bisulfite DNA sequencing and mRNA levels by quantitative RT-PCR reveled that while gene silencing and DNA methylation were established efficiently at these loci upon expression of LSH in the control *Lsh*−/− shNS MEFs, both *de novo* DNA methylation (Figure 6C) and transcriptional silencing (Figure 5D) were impaired in DNMT3B knockdown sh2D3B-LSH MEFs. These experiments demonstrate that the cellular concentration of DNMT3B is critical for LSH-mediated gene silencing and *de novo* DNA methylation in somatic cells.

## DISCUSSION

It is well established that early embryogenesis is orchestrated by morphogen gradients, signaling pathways and strong transcription factors, which shape the repertoire of gene expression in differentiating cells. A number of studies have linked chromatin modifications and DNA methylation to developmental signaling (36) and DNA methylation reprogramming to major developmental transitions from toti- and pluripotency to developmentally restricted cell fates (2,3). Given this knowledge, we asked in this study whether or not *de novo* DNA methylation can occur in somatic cells in the absence of exogenous signals.

The *Lsh*−/− MEFs are an excellent model to address this question as it is assumed that in the absence of LSH-dependent chromatin remodeling a proportion of developmentally programmed *de novo* DNA methylation events have failed to occur in embryos from which these cells are derived (18,19). Thus, we aimed to determine if restoring the LSH expression in *Lsh*−/− MEFs would result in patterns of DNA methylation that resemble those of their wild-type counterparts. This is an important question as recently mutations in LSH were found in a subset of patients with ICF (Immunodeficiency, Centromeric instability and Facial anomalies) syndrome (37), a human disease most commonly associated with congenital mutations in *DNMT3B* (38,39). Thus, addressing the reversibility of DNMT3B and LSH deficiency-caused phenotypes in differentiated cells is instrumental to motivate further research into improved therapies for ICF syndrome patients.

Surprisingly, we found that the expression of LSH in *Lsh*−/− MEFs grown in culture for many cell generations substantially reestablishes DNA methylation in the absence of additional signals. We also found that LSH supports *de novo* DNA methylation by both DNMT3 paralogues, DNMT3A and 3B, and that an appropriate cellular concentration of DNMT3B as well as the ability of LSH to bind and hydrolyse ATP are critical for efficient *de novo* DNA methylation at unique and repetitive sequences. The latter is in agreement with recent reports (25,40) and strongly suggests that chromatin remodeling by LSH is necessary for its function in DNA methylation and double-strand break repair.

*De novo* DNMTs display high affinity for chromatin (34,41) and complex binding patterns across the genome, which are determined by the CpG density and histone modifications (15). In contrast, LSH displays low affinity for DNA/chromatin and a highly dynamic behavior (42). Thus it is conceivable that the patterns of *de novo* DNMT occupancy remain unchanged in the *Lsh*−/− MEFs and that these patterns and the overall residence time of DNMTs on chromatin, which might be higher on stably positioned nucleosomes, could determine the sites where chromatin remodeling by LSH is needed for efficient *de novo* DNA methylation. Such model would explain the accurate, yet incomplete, reestablishment of DNA methylation patterns in the *Lsh*−/− MEFs upon LSH re-expression. In agreement with this, we demonstrate that that the appropriate concentration of DNMT3B is required for LSH-dependent gene silencing and *de novo* DNA methylation. Thus the partial knockdown of DNMT3B in the *Lsh*−/− MEFs prior to LSH expression impaired both silencing and *de novo* methylation of LSH-dependent loci. Since the expression of DNMT3A and 3B, but not DNMT1, is downregulated upon differentiation of embryonic cells (9,19), the low concentration of *de novo* enzymes as well as the absence of DNMT3L may limit the extent of LSH-mediated reestablishment of DNA methylation in somatic cells. We attempted to overexpress either DNMT3A and 3B in *Lsh*−/− LSH MEFs, but were unable to isolate stable clones, which could indicate that somatic cells do not tolerate elevated expression of *de novo* DNMTs.

Importantly, the LSH-induced *de novo* DNA methylation in *Lsh*−/− cells occurred at inactive as well as active, H3K4me3-marked, promoters resulting in stable silencing of associated genes and transposons, as demonstrated here for the *Rhox* loci and IAPs. This raises the question of how LSH and *de novo* DNMTs are recruited to active promoters. As the binding of DNMT3A and 3B to the N-terminus of H3 is inhibited by the presence H3K4 di- and trimethylation (34,41), it seems unlikely that the recruitment of LSH to active promoters is directed by either of the two DNMT3 paralogues. Addressing this question requires further studies in an inducible system that would enable controlled re-expression of LSH in the *Lsh*−/− MEFs and detailed investigation of chromatin dynamics, transcriptional silencing and the establishment of *de novo* DNA methylation following LSH induction. In parallel, a better characterization of LSH-associated proteins may also shed light on the mechanisms that govern the establishment of LSH-dependent heterochromatin.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Kathrin Muegge (NCI, Frederick) for providing the *Lsh*−/− and control wild-type MEFs; the staff of Edinburgh Genomics and in particular Karim Gharbi for assistance and sequencing of MeDIP libraries and the members of Bird and Stancheva lab for stimulating discussions.

## FUNDING

Cancer Research UK [C7215/A8983, C7215/A9218]; Wellcome Trust Centre for Cell Biology [092076]. N.T. is supported by a Darwin Trust PhD scholarship. The open access publication charge for this paper has been waived by Oxford University Press - *NAR* Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.
*Conflict of interest statement.* None declared.

## REFERENCES

1. Suzuki,M.M. and Bird,A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.*, **9**, 465–476.
2. Smith,Z.D. and Meissner,A. (2013) DNA methylation: roles in mammalian development. *Nat. Rev. Genet.*, **14**, 204–220.
3. Reik,W. (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, **447**, 425–432.
4. Goll,M.G. and Bestor,T.H. (2005) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.*, **74**, 481–514.
5. Smith,Z.D., Chan,M.M., Mikkelsen,T.S., Gu,H., Gnirke,A., Regev,A. and Meissner,A. (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, **484**, 339–344.
6. Borgel,J., Guibert,S., Li,Y., Chiba,H., Schubeler,D., Sasaki,H., Forne,T. and Weber,M. (2010) Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet.*, **42**, 1093–1100.
7. Li,E., Bestor,T.H. and Jaenisch,R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915–926.
8. Okano,M., Bell,D.W., Haber,D.A. and Li,E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.
9. Gendrel,A.V., Apedaile,A., Coker,H., Termanis,A., Zvetkova,I., Godwin,J., Tang,Y.A., Huntley,D., Montana,G., Taylor,S. *et al.* (2012) Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev. Cell*, **23**, 265–279.
10. Oda,M., Yamagiwa,A., Yamamoto,S., Nakayama,T., Tsumura,A., Sasaki,H., Nakao,K., Li,E. and Okano,M. (2006) DNA methylation regulates long-range gene silencing of an X-linked homeobox gene cluster in a lineage-specific manner. *Genes Dev.*, **20**, 3382–3394.
11. Leitch,H.G., McEwen,K.R., Turp,A., Encheva,V., Carroll,T., Grabole,N., Mansfield,W., Nashun,B., Knezovich,J.G., Smith,A. *et al.* (2013) Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.*, **20**, 311–316.
12. Habibi,E., Brinkman,A.B., Arand,J., Kroeze,L.I., Kerstens,H.H., Matarese,F., Lepikhov,K., Gut,M., Brun-Heath,I., Hubner,N.C. *et al.* (2013) Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell*, **13**, 360–369.
13. Ficz,G., Hore,T.A., Santos,F., Lee,H.J., Dean,W., Arand,J., Krueger,F., Oxley,D., Paul,Y.L., Walter,J. *et al.* (2013) FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell*, **13**, 351–359.
14. Felle,M., Hoffmeister,H., Rothammer,J., Fuchs,A., Exler,J.H. and Langst,G. (2011) Nucleosomes protect DNA from DNA methylation in vivo and in vitro. *Nucleic Acids Res.*, **39**, 6956–6969.
15. Baubec,T., Colombo,D.F., Wirbelauer,C., Schmidt,J., Burger,L., Krebs,A.R., Akalin,A. and Schubeler,D. (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*, **520**, 243–247.
16. Dennis,K., Fan,T., Geiman,T., Yan,Q. and Muegge,K. (2001) Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev.*, **15**, 2940–2944.
17. Gibbons,R.J., McDowell,T.L., Raman,S., O'Rourke,D.M., Garrick,D., Ayyub,H. and Higgs,D.R. (2000) Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat. Genet.*, **24**, 368–371.
18. Yu,W., McIntosh,C., Lister,R., Zhu,I., Han,Y., Ren,J., Landsman,D., Lee,E., Briones,V., Terashima,M. *et al.* (2014) Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. *Genome Res.*, **24**, 1613–1623.
19. Myant,K., Termanis,A., Sundaram,A.Y., Boe,T., Li,C., Merusi,C., Burrage,J., de Las Heras,J.I. and Stancheva,I. (2011) LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. *Genome Res.*, **21**, 83–94.
20. Myant,K. and Stancheva,I. (2008) LSH cooperates with DNA methyltransferases to repress transcription. *Mol. Cell. Biol.*, **28**, 215–226.
21. Bock,C., Reither,S., Mikeska,T., Paulsen,M., Walter,J. and Lengauer,T. (2005) BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics*, **21**, 4067–4068.
22. Ramsahoye,B.H. (2002) Measurement of genome wide DNA methylation by reversed-phase high-performance liquid chromatography. *Methods*, **27**, 156–161.
23. Pfaffl,M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.*, **29**, e45.
24. Mohn,F., Weber,M., Schubeler,D. and Roloff,T.C. (2009) Methylated DNA immunoprecipitation (MeDIP). *Methods Mol Biol.*, **507**, 55–64.
25. Burrage,J., Termanis,A., Geissner,A., Myant,K., Gordon,K. and Stancheva,I. (2012) The SNF2 family ATPase LSH promotes phosphorylation of H2AX and efficient repair of DNA double-strand breaks in mammalian cells. *J. Cell Sci.*, **125**, 5524–5534.
26. Muegge,K. (2005) Lsh, a guardian of heterochromatin at repeat elements. *Biochem. Cell Biol.*, **83**, 548–554.
27. Yan,Q., Huang,J., Fan,T., Zhu,H. and Muegge,K. (2003) Lsh, a modulator of CpG methylation, is crucial for normal histone methylation. *EMBO J.*, **22**, 5154–5162.
28. Dunican,D.S., Cruickshanks,H.A., Suzuki,M., Semple,C.A., Davey,T., Arceci,R.J., Greally,J., Adams,I.R. and Meehan,R.R. (2013) Lsh regulates LTR retrotransposon repression independently of Dnmt3b function. *Genome Biol.*, **14**, R146.
29. Lee,D.S., Shin,J.Y., Tonge,P.D., Puri,M.C., Lee,S., Park,H., Lee,W.C., Hussein,S.M., Bleazard,T., Yun,J.Y. *et al.* (2014) An epigenomic roadmap to induced pluripotency reveals DNA methylation as a reprogramming modulator. *Nat. Commun.*, **5**, 5619.
30. Lee,S.E., Lee,S.Y. and Lee,K.A. (2013) Rhox in mammalian reproduction and development. *Clin. Exp. Reprod. Med.*, **40**, 107–114.
31. Santos-Rosa,H., Schneider,R., Bannister,A.J., Sherriff,J., Bernstein,B.E., Emre,N.C., Schreiber,S.L., Mellor,J. and Kouzarides,T. (2002) Active genes are tri-methylated at K4 of histone H3. *Nature*, **419**, 407–411.
32. Mikkelsen,T.S., Ku,M., Jaffe,D.B., Issac,B., Lieberman,E., Giannoukos,G., Alvarez,P., Brockman,W., Kim,T.K., Koche,R.P. *et al.* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, **448**, 553–560.
33. Ooi,S.K., Qiu,C., Bernstein,E., Li,K., Jia,D., Yang,Z., Erdjument-Bromage,H., Tempst,P., Lin,S.P., Allis,C.D. *et al.* (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, **448**, 714–717.
34. Otani,J., Nankumo,T., Arita,K., Inamoto,S., Ariyoshi,M. and Shirakawa,M. (2009) Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep.*, **10**, 1235–1241.
35. Zhang,Y., Jurkowska,R., Soeroes,S., Rajavelu,A., Dhayalan,A., Bock,I., Rathert,P., Brandt,O., Reinhardt,R., Fischle,W. *et al.* (2010) Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res.*, **38**, 4246–4253.

36. Albert,M. and Peters,A.H. (2009) Genetic and epigenetic control of early mouse development. *Curr. Opin. Genet. Dev.*, **19**, 113–121.
37. Thijssen,P.E., Ito,Y., Grillo,G., Wang,J., Velasco,G., Nitta,H., Unoki,M., Yoshihara,M., Suyama,M., Sun,Y. *et al.* (2015) Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. *Nat. Commun.*, **6**, 7870.
38. Hansen,R.S., Wijmenga,C., Luo,P., Stanek,A.M., Canfield,T.K., Weemaes,C.M. and Gartler,S.M. (1999) The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 14412–14417.
39. Wijmenga,C., Hansen,R.S., Gimelli,G., Bjorck,E.J., Davies,E.G., Valentine,D., Belohradsky,B.H., van Dongen,J.J., Smeets,D.F., van den Heuvel,L.P. *et al.* (2000) Genetic variation in ICF syndrome: evidence for genetic heterogeneity. *Hum. Mutat.*, **16**, 509–517.
40. Ren,J., Briones,V., Barbour,S., Yu,W., Han,Y., Terashima,M. and Muegge,K. (2015) The ATP binding site of the chromatin remodeling homolog Lsh is required for nucleosome density and de novo DNA methylation at repeat sequences. *Nucleic Acids Res.*, **43**, 1444–1455.
41. Noh,K.M., Wang,H., Kim,H.R., Wenderski,W., Fang,F., Li,C.H., Dewell,S., Hughes,S.H., Melnick,A.M., Patel,D.J. *et al.* (2015) Engineering of a Histone-Recognition Domain in Dnmt3a Alters the Epigenetic Landscape and Phenotypic Features of Mouse ESCs. *Mol. Cell*, **59**, 89–103.
42. Lungu,C., Muegge,K., Jeltsch,A. and Jurkowska,R.Z. (2015) An ATPase-deficient variant of the SNF2 family member HELLS shows altered dynamics at pericentromeric heterochromatin. *J. Mol. Biol.*, **427**, 1903–1915.

## References


Alvaro, D., Lisby, M., & Rothstein, R. (2007). Genome-wide analysis of Rad52 foci reveals diverse mechanisms impacting recombination. *PLoS Genetics*, *3*(12), 2439–2449. http://doi.org/10.1371/journal.pgen.0030228

Ambrosi, C., Manzo, M., & Baubec, T. (2017). Dynamics and Context-Dependent Roles of DNA Methylation. *Journal of Molecular Biology*, *429*(10), 1459–1475. http://doi.org/10.1016/j.jmb.2017.02.008

Amir, R. E., Van den Veyver, I. . B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics*, *23*(october), 185–188. http://doi.org/10.1038/13810

Antequera, F., & Bird, A. (1999). CpG islands as genomic footprints of promoters that are associated with replication origins. *Current Biology*, *9*(17), 661–667. http://doi.org/10.1016/S0960-9822(99)80418-7

Antequera, F., & Bird, a. (1993). Number of CpG islands and genes in human and mouse. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(24), 11995–11999. http://doi.org/10.1073/pnas.90.24.11995

Athanasiadou, R., de Sousa, D., Myant, K., Merusi, C., Stancheva, I., & Bird, A. (2010). Targeting of De novo DNA methylation throughout the Oct-4 gene regulatory region in differentiating embryonic stem cells. *PLoS ONE*, *5*(4), 1–10. http://doi.org/10.1371/journal.pone.0009937

Auclair, G., Borgel, J., Sanz, L. A., Vallet, J., Guibert, S., Dumas, M. 3, … Weber, M. (2016). EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos Running title: EHMT2 regulates DNA methylation in mouse embryos, 192–202. http://doi.org/10.1101/gr.198291.115

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381–395. http://doi.org/10.1038/cr.2011.22

Basenko, E. Y., Kamei, M., Ji, L., Schmitz, R. J., & Lewis, Z. A. (2016). The LSH/DDM1 Homolog MUS-30 Is Required for Genome Stability, but Not for DNA Methylation in Neurospora crassa. *PLOS Genetics*, *12*(1), e1005790. http://doi.org/10.1371/journal.pgen.1005790

Baubec, T., Colombo, D. F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A. R., … Schübeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*. http://doi.org/10.1038/nature14176

Becker, P. B., & Hörz, W. (2002). ATP-dependent nucleosome remodeling. *Annual Review of Biochemistry*, *71*, 247–73. http://doi.org/10.1146/annurev.biochem.71.110601.135400

Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J., & Woodcock, C. L. (1998). Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proceedings of the National Academy of Sciences*, *95*(24), 14173–14178. http://doi.org/10.1073/pnas.95.24.14173

Bestor, T. H. (2000). The DNA methyltransferases of mammals. *Human Molecular Genetics*, *9*(16), 2395–2402. http://doi.org/10.1093/hmg/9.16.2395

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes & Development*, *16*(1), 6–21. http://doi.org/10.1101/gad.947102

Bird, A. (2007). Perceptions of epigenetics. *Nature*, *447*, 396–398. http://doi.org/10.1038/nature05913

Bird, A. P. (1986). CpG-Rich islands and the function of DNA methylation. *Nature*, *321*(6067), 209–213. http://doi.org/10.1038/321209a0

Bird, A., Taggart, M., Frommer, M., Miller, O. J., & Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, *40*(1), 91–99. http://doi.org/10.1016/0092-8674(85)90312-5

Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L. P., … Klose, R. J. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell*, *157*(6), 1445–1459. http://doi.org/10.1016/j.cell.2014.05.004

Boiani, M., & Schöler, H. R. (2005). Regulatory networks in embryo-derived pluripotent stem cells. *Nature Reviews Molecular Cell Biology*, *6*(11), 872–884. http://doi.org/10.1038/nrm1744

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., … Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. *Nature Genetics*, *42*(12), 1093–1100. http://doi.org/10.1038/ng.708

Bork, S., Pfister, S., Horn, P., Korn, B., Anthony, D., & Wagner, W. (2010). DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells. *Aging Cell*, (9), 54–63. http://doi.org/10.1111/j.1474-9726.2009.00535.x

Bowman, G. D. (2010). Mechanisms of ATP-dependent nucleosome sliding. *Current Opinion in Structural Biology*, *20*(1), 73–81. http://doi.org/10.1016/j.sbi.2009.12.002

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., … Jaenisch, R. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, *441*(7091), 349–353. http://doi.org/10.1038/nature04733

Briones, V., & Muegge, K. (2012). The ghosts in the machine: DNA methylation and the mystery of differentiation. *Biochimica et Biophysica Acta*, *1819*(7), 757–62. http://doi.org/10.1016/j.bbagrm.2012.02.013

Brzeski, J., & Jerzmanowski, A. (2003). Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors. *Journal of Biological Chemistry*, *278*(2), 823–828. http://doi.org/10.1074/jbc.M209260200

Burkhart, D. L., & Sage, J. (2008). Cellular mechanisms of tumour suppression by the retinoblastoma gene. *Nature Reviews Cancer*, *8*(9), 671–682. http://doi.org/10.1038/nrc2399

Burrage, J., Termanis, A., Geissner, A., Myant, K., Gordon, K., & Stancheva, I. (2012). The SNF2 family ATPase LSH promotes phosphorylation of H2AX and efficient repair of DNA double-strand breaks in mammalian cells. *Journal of Cell Science*, *125*(Pt 22), 5524–34. http://doi.org/10.1242/jcs.111252

Cairns, B. R. (2007). Chromatin remodeling: insights and intrigue from single-molecule studies. *Nat Struct Mol Biol*, *14*(11), 989–996. http://doi.org/10.1038/nsmb1333.Chromatin

Cano-rodriguez, D., Gjaltema, R. A. F., Jilderda, L. J., Jellema, P., & Dokter-, J. (2016). Writing of H3K4Me3 overcomes epigenetic silencing in a sustained , but context-dependent manner. *Nature Communications*, *7*, 1–11. http://doi.org/10.1038/ncomms12284

Catez, F., Yang, H., Tracey, K. J., Reeves, R., Misteli, T., & Bustin, M. (2004). Network of dynamic interactions between histone H1 and high-mobility-group proteins in chromatin. *Molecular and Cellular Biology*, *24*(10), 4321–8. http://doi.org/10.1128/MCB.24.10.4321

Cedar, H., & Almouzni, G. (2016). Maintenance of Epigenetic Information. *Cold Spring Harb Perspect Biol*, *8*.

Chapman, V., Forrester, L., Sanford, J., Hastie, N., & Rossant, J. (1984). Cell lineage-specific undermethylation of mouse repetitive DNA. *Nature*, *307*(5948), 284–286. http://doi.org/10.1038/307284a0

Chen, T., Ueda, Y., Dodge, J. E., Wang, Z., & Li, E. (2003). Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Molecular and Cellular Biology*, *23*(16), 5594–5605. http://doi.org/10.1128/MCB.23.16.5594

Chodavarapu, R. K., Feng, S., Bernatavichute, Y. V, Stroud, H., Yu, Y., Hetzel, J., … Jacobsen, S. E. (2010). Relationship between nucleosome positioning and DNA methylation. *Nature*, *466*(7304), 388–392. http://doi.org/10.1038/nature09147.Relationship

Clapier, C. R., & Cairns, B. R. (2009). The Biology of Chromatin Remodeling Complexes. *Annual Review of Biochemistry*, *78*(1), 273–304. http://doi.org/10.1146/annurev.biochem.77.062706.153223

Clouaire, T., Webb, S., & Bird, A. (2014). Cfp1 is required for gene expression-dependent H3K4 trimethylation and H3K9 acetylation in embryonic stem cells. *Genome Biology*, *15*(9), 451. http://doi.org/10.1186/s13059-014-0451-x

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., … Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes and Development*, *26*(15), 1714–1728. http://doi.org/10.1101/gad.194209.112

Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., … Jacobsen, S. E. (2008). Shotgun bisulfite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, *452*(7184), 215–219. http://doi.org/10.1038/nature06745

Corona, D. F. V, La, G., Clapier, C. R., Bonte, E. J., Ferrari, S., Tamkun, J. W., & Becker, P. B. (1999). ISWI Is an ATP-Dependent Nucleosome Remodeling Factor. *Molecular Cell*, *3*, 239–245.

Cruickshanks, H. A., McBryan, T., Nelson, D. M., VanderKraats, N. D., Shah, P. P., Tuyn, J. van, … Adams, P. D. (2013). Senescent cells harbour features of the cancer epigenome. *Nat Cell Biol.*, *134*(22), 1–10. http://doi.org/10.1021/ja303183z.Aqueous

Curie, I., & Paris, U. (2011). The Polycomb Complex PRC2 and its Mark in Life. *Nature*, *469*(7330), 343–349. http://doi.org/10.1038/nature09784.

De Greef, J. C., Wang, J., Balog, J., Den Dunnen, J. T., Frants, R. R., Straasheijm, K. R., … Van Der Maarel, S. M. (2011). Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. *American Journal of Human Genetics*, *88*(6), 796–804. http://doi.org/10.1016/j.ajhg.2011.04.018

De La Fuente, R., Baumann, C., Fan, T., Schmidtmann, A., Dobrinski, I., & Muegge, K. (2006). Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells. *Nature Cell Biology*, *8*(12), 1448–54. http://doi.org/10.1038/ncb1513

Dennis, K., Fan, T., Geiman, T., Yan, Q., & Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes & Development*, *15*(22), 2940–4. http://doi.org/10.1101/gad.929101

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R. Z., Ragozin, S., & Jeltsch, A. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *Journal of Biological Chemistry*, *285*(34), 26114–26120. http://doi.org/10.1074/jbc.M109.089433

Dorigo, B., Schalch, T., Bystricky, K., & Richmond, T. J. (2003). Chromatin fiber folding: Requirement for the histone H4 N-terminal tail. *Journal of Molecular Biology*, *327*(1), 85–96. http://doi.org/10.1016/S0022-2836(03)00025-1

Du, J., Johnson, L. M., Jacobsen, S. E., & Patel, D. J. (2015). DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol*, *16*(9), 519–532. http://doi.org/10.1038/nrm4043.DNA

Dunican, D. S., Cruickshanks, H. a, Suzuki, M., Semple, C. a, Davey, T., Arceci, R. J., … Meehan, R. R. (2013). Lsh regulates LTR retrotransposon repression independently of Dnmt3b function. *Genome Biology*, *14*(12), R146. http://doi.org/10.1186/gb-2013-14-12-r146

Duymich, C. E., Charlet, J., Yang, X., Jones, P. A., & Liang, G. (2016). DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. *Nature Communications*, *7*, 11453. http://doi.org/10.1038/ncomms11453

Ehrlich, M., Gama-Sosa, M. A., Huang, L. H., Midgett, R. M., Kuo, K. C., Mccune, R. A., & Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Research*, *10*(8), 2709–2721. http://doi.org/10.1093/nar/10.8.2709

Eisen, J. A., Sweder, K. S., & Hanawalt, P. C. (1995). Evolution of the SNF2 family of proteins : subfamilies with distinct sequences and functions. *Nucleic Acids Research, 23*(14), 2715–2723.

Epsztejn-Litman, S., Feldman, N., Abu-remaileh, M., Shufaro, Y., Ueda, J., Deplus, R., … Bergman, Y. (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nature Structural & Molecular Biology*, *15*(11), 1176–1183. http://doi.org/10.1038/nsmb.1476.De

Fan, J. Y., Rangasamy, D., Luger, K., & Tremethick, D. J. (2004). H2A.Z alters the nucleosome surface to promote HP1α-mediated chromatin fiber folding. *Molecular Cell*, *16*(4), 655–661. http://doi.org/10.1016/j.molcel.2004.10.023

Fan, T., Hagan, J. P., Kozlov, S. V, Stewart, C. L., & Muegge, K. (2005). Lsh controls silencing of the imprinted Cdkn1c gene. *Development (Cambridge, England)*, *132*(4), 635–44. http://doi.org/10.1242/dev.01612

Fan, Y., Nikitina, T., Zhao, J., Fleury, T. J., Bhattacharyya, R., Bouhassira, E. E., … Skoultchi, A. I. (2005). Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell*, *123*(7), 1199–1212. http://doi.org/10.1016/j.cell.2005.10.028

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., … Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nature Cell Biology*, *8*(2), 188–194. http://doi.org/10.1038/ncb1353

Ferry, L., Fournier, A., Tsusaka, T., Adelmant, G., Shimazu, T., Matano, S., … Defossez, P.-A. (2017). Methylation of DNA Ligase 1 by G9a/GLP Recruits UHRF1 to Replicating DNA and Regulates DNA Methylation. *Molecular Cell*, *7*(0), 12464. http://doi.org/10.1016/j.molcel.2017.07.012

Filion, G. J., Bemmel, J. G. Van, Braunschweig, U., & Talhout, W. (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Ce*, *143*(2), 212–224. http://doi.org/10.1016/j.cell.2010.09.009.Systematic

Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D., & Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by polycomb and HP1 chromodomains. *Genes and Development*, *17*(15), 1870–1881. http://doi.org/10.1101/gad.1110503

Flaus, A., Martin, D. M. a, Barton, G. J., & Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Research*, *34*(10), 2887–905. http://doi.org/10.1093/nar/gkl295

Flaus, A., & Owen-Hughes, T. (2011). Mechanisms for ATP-dependent chromatin remodeling: the means to the end. *FEBS Journal*, *278*(19), 3579–3595. http://doi.org/10.1111/j.1742-4658.2011.08281.x

Franzen, J., Zirkel, A., Blake, J., Benes, V., Papantonis, A., & Wagner, W. (2017). Senescence-associated DNA methylation is stochastically acquired in subpopulations of mesenchymal stem cells. *Aging Cell*, (16), 183–191. http://doi.org/10.1111/acel.12544

Garrick, D., Gibbons, R. J., McDowell, T. L., Raman, S., O'Rourke, D. M., Ayyub, H., & Higgs, D. R. (2000). Mutations in the human ATRX gene, encoding a putative transcription regulator, cause diverse changes in the pattern of DNA methylation. *Blood Cells Molecules and Diseases*, *26*(5), 515–516.

Gates, L. A., Shi, J., Rohira, A. D., Feng, Q., Zhu, B., Bedford, M. T., … O'Malley, B. W. (2017). Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. *Journal of Biological Chemistry*, (1), jbc.M117.802074. http://doi.org/10.1074/jbc.M117.802074

Geiman, T. M., Durum, S. K., & Muegge, K. (1998). Characterization of Gene Expression , Genomic Structure , and Chromosomal Localization of Hells ( Lsh ). *Genomics*, *54*, 477–483.

Geiman, T. M., & Muegge, K. (2010). DNA methylation in early development. *Molecular Reproduction and Development*, *77*(2), 105–13. http://doi.org/10.1002/mrd.21118

Geiman, T. M., Tessarollo, L., Anver, M. R., Kopp, J. B., Ward, J. M., & Muegge, K. (2001). Lsh, a SNF2 family member, is required for normal murine development. *Biochimica et Biophysica Acta*, *1526*(2), 211–20. http://www.ncbi.nlm.nih.gov/pubmed/11325543

Gopalakrishnan, S., Emburgh, B. O. Van, & Robertson, K. D. (2008). DNA methylation in development and human disease. *Mutat Res*, *647*, 30–38. http://doi.org/10.1016/j.mrfmmm.2008.08.006.DNA

Gottlicher, M., Minucci, S., Zhu, P., Kra, O. H., Schimpf, A., Giavara, S., … Heinzel, T. (2001). Valproic acid defines a novel class of HDAC inhibitors inducing differentiation of transformed cells. *EMBO Journal*, *20*(24), 6969–6978. http://doi.org/10.1093/emboj/20.24.6969

Hansen, J. C., Ghosh, R. P., & Woodcock, C. L. (2010). Binding of the Rett syndrome protein, MeCP2, to methylated and unmethylated DNA and chromatin. *IUBMB Life*, *62*(10), 732–738. http://doi.org/10.1002/iub.386

Hervouet, E., Peixoto, P., Delage-mourroux, R., Boyer-guittaut, M., & Cartron, P. (2018). Specific or not specific recruitment of DNMTs for DNA methylation , an epigenetic dilemma, 1–18. http://doi.org/10.1186/s13148-018-0450-y

Ho, L., & Crabtree, G. R. (2010). Chromatin remodelling during development. *Nature*, *463*(7280), 1–11. http://doi.org/10.1038/nature08911.Chromatin

Hoffmeister, H., Fuchs, A., Erdel, F., Pinz, S., Gr, R., Bruckmann, A., … Dna, H. (2017). CHD3 and CHD4 form distinct NuRD complexes with different yet overlapping functionality Gernot L angst. *Nucleic Acids Research*, *45*(18), 10534–10554. http://doi.org/10.1093/nar/gkx711

Huang, J., Fan, T., Yan, Q., Zhu, H., Fox, S., Issaq, H. J., … Muegge, K. (2004). Lsh, an epigenetic guardian of repetitive elements. *Nucleic Acids Research*, *32*(17), 5019–28. http://doi.org/10.1093/nar/gkh821

Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R. W., James, K. D., Turner, D. J., … Bird, A. P. (2010). Orphan CpG Islands Identify numerous conserved promoters in the mammalian genome. *PLoS Genetics*, *6*(9). http://doi.org/10.1371/journal.pgen.1001134

Indra, A. K., Warot, X., Brocard, J., Bornert, J. M., Xiao, J. H., Chambon, P., & Metzger, D. (1999). Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: Comparison of the recombinase activity of the tamoxifen-inducible Cre-ER(T) and Cre-ER(T2) recombinases. *Nucleic Acids Research*, *27*(22), 4324–4327. http://doi.org/10.1093/nar/27.22.4324

Jarvis, C. D., Geiman, T., Vila-Storm, M. P., Osipovich, O., Akella, U., Candeias, S., … Muegge, K. (1996). A novel putative helicase produced in early murine lymphocytes. *Gene*, *169*(2), 203–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8647447

Jeddeloh, J. A., Stokes, T. L., & Richards, E. J. (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nature Genetics*, *22*(1), 94–97. http://doi.org/10.1038/8803

Jeddeloh, J., & Bender, J. (1998). The DNA methylation locusDDM1 is required for maintenance of gene silencing in Arabidopsis. *Genes & Development*, *12*, 1714–1725. http://doi.org/10.1101/gad.12.11.1714

Jenness, C., Giunta, S., Müller, M. M., Kimura, H., Muir, T. W., & Funabiki, H. (2018). HELLS and CDCA7 comprise a bipartite nucleosome remodeling complex defective in ICF syndrome. *Proceedings of the National Academy of Sciences*, 201717509. http://doi.org/10.1073/pnas.1717509115

Jermann, P., Hoerner, L., Burger, L., & Schübeler, D. (2014). Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(33), E3415-21. http://doi.org/10.1073/pnas.1400672111

Jiang, Y. L., Rigolet, M., Bourc'his, D., Nigon, F., Bokesoy, I., Fryns, J. P., … Viegas-Péquignot, E. (2005). DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. *Human Mutation*, *25*(1), 56–63. http://doi.org/10.1002/humu.20113

Jones, P. A. (2012). Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, *13*(7), 484–492. http://doi.org/10.1038/nrg3230

Jones, P. A., & Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. *Nature Reviews Genetics*, *3*(6), 415–428. http://doi.org/10.1038/nrg816

Jones, P. a, & Baylin, S. B. (2007). The epigenomics of cancer. *Cell*, *128*(4), 683–92. http://doi.org/10.1016/j.cell.2007.01.029

Jurkowski, T. P., Shanmugam, R., Helm, M., & Jeltsch, A. (2012). Mapping the tRNA Binding Site on the Surface of Human DNMT2 Methyltransferase. *Biochemistry*, (51), 4438–4444. http://doi.org/10.1021/bi3002659

Kelly, T. K., Miranda, T. B., Liang, G., Berman, B. P., Lin, J. C., Tanay, A., & Jones, P. A. (2010). H2A.Z Maintenance During Mitosis Reveals Nucleosome Shifting on Mitotically Silenced Genes. *Mol Cell*, *39*(6), 901–911. http://doi.org/10.1016/j.molcel.2010.08.026.H2A.Z

Keyes, W. M., Pecoraro, M., Aranda, V., Vernersson-Lindahl, E., Li, W., Vogel, H., … Mills, A. a. (2011). ΔNp63α is an oncogene that targets chromatin remodeler Lsh to drive skin stem cell proliferation and tumorigenesis. *Cell Stem Cell*, *8*(2), 164–76. http://doi.org/10.1016/j.stem.2010.12.009

Kim, S.-K., Suh, M. R., Yoon, H. S., Lee, J. B., Oh, S. K., Moon, S. Y., … Kim, K.-S. (2005). Identification of Developmental Pluripotency Associated 5 Expression in Human Pluripotent Stem Cells. *Stem Cells*, *23*(4), 458–462. http://doi.org/10.1634/stemcells.2004-0245

Klose, R. J., & Bird, A. P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences*, *31*(2), 89–97. http://doi.org/10.1016/j.tibs.2005.12.008

Koch, C. M., Reck, K., Shao, K., Lin, Q., Joussen, S., Ziegler, P., … Tomo, S. (2013). Pluripotent stem cells escape from senescence- associated DNA methylation changes. *Genome Research*, (23), 248–259. http://doi.org/10.1101/gr.141945.112.248

Kornberg, R. (1974). Chromatin Structure : A Repeating Unit of Histones and DNA Chromatin structure is based on a repeating unit of eight. *Science*, *184*, 868–871.

Krisnamurti, D. G. B., Louisa, M., Anggraeni, E., & Wanandi, S. I. (2016). Drug Efflux Transporters Are Overexpressed in Short-Term Tamoxifen-Induced MCF7 Breast Cancer Cells. *Advances in Pharmacological Sciences*, 6–11. http://doi.org/10.1155/2016/6702424

Kuzmichev, A., Nishioka, K., Edrjument-Bromage, H., Tempst, P., & Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes & Development*, (16), 2893–2905. http://doi.org/10.1101/gad.1035902.repression

Längst, G., & Becker, P. B. (2004). Nucleosome remodeling: one mechanism, many phenomena? *Biochimica et Biophysica Acta*, *1677*(1–3), 58–63. http://doi.org/10.1016/j.bbaexp.2003.10.011

Laurent, B. C., Treitel, M. a, & Carlson, M. (1991). Functional interdependence of the yeast SNF2, SNF5, and SNF6 proteins in transcriptional activation. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(7), 2687–2691. http://doi.org/10.1073/pnas.88.7.2687

Laurent, L., Wong, E., Li, G., Laurent, L., Wong, E., Li, G., … Sung, K. (2010). Dynamic changes in the human methylome during differentiation. *Genome Research*, (20), 320–331. http://doi.org/10.1101/gr.101907.109

Lee, D. W., Zhang, K., Ning, Z., Raabe, E. H., Tintner, S., Wieland, R., … Arceci, R. J. (2000). Proliferation-associated SNF2-like Gene ( PASG ): A SNF2 Family Member Altered in Leukemia. *Cancer Research*, (60), 3612–3622.

Lee, S.-M., Lee, J., Noh, K.-M., Choi, W.-Y., Jeon, S., Oh, G. T., … Kim, Y.-J. (2017). Intragenic CpG islands play important roles in bivalent chromatin assembly of developmental genes. *Proceedings of the National Academy of Sciences*, *114*(10), E1885–E1894. http://doi.org/10.1073/pnas.1613300114

Leeb, M., & Wutz, A. (2012). Establishment of epigenetic patterns in development. *Chromosoma*, *121*(3), 251–62. http://doi.org/10.1007/s00412-012-0365-x

Lehnertz, B., Ueda, Y., Derijck, A. A. H. A., Braunschweig, U., Perez-burgos, L., Kubicek, S., … Peters, A. H. F. M. (2003). Suv39h -Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Current Biology*, *13*, 1192–1200. http://doi.org/10.1016/S

Leonhardt, H., Page, A. W., Weier, H. U., & Bestor, T. H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*, *71*(5), 865–873. http://doi.org/10.1016/0092-8674(92)90561-P

Li, E., Bestor, T. H., & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, *69*(6), 915–926. http://doi.org/10.1016/0092-8674(92)90611-F

Lister, R., Pelizzola, M., Dowen, R., Hawkins, Rd., Hon, G., Nery, J., … Ecker, J. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–322. http://doi.org/10.1038/nature08514.Human

Loh, Y. H., Wu, Q., Chew, J. L., Vega, V. B., Zhang, W., Chen, X., … Ng, H. H. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genetics*, *38*(4), 431–440. http://doi.org/10.1038/ng1760

Long, H. K., Blackledge, N. P., & Klose, R. J. (2013). ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochemical Society Transactions*, *41*(3), 727–740. http://doi.org/10.1042/BST20130028

Luger, K. (2003). Structure and dynamic behavior of nucleosomes. *Current Opinion in Genetics and Development*, *13*(2), 127–135. http://doi.org/10.1016/S0959-437X(03)00026-1

Luger, K., & Hansen, J. C. (2005). Nucleosome and chromatin fiber dynamics. *Current Opinion in Structural Biology*, *15*(2), 188–196. http://doi.org/10.1016/j.sbi.2005.03.006

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, *389*(6648), 251–260. http://doi.org/10.1038/38444

Luger, K., & Richmond, T. J. (1998). The histone tails of the nucleosome. *Current Opinion in Genetics and Development*, *8*(2), 140–146. http://doi.org/10.1016/S0959-437X(98)80134-2

Lungu, C., Muegge, K., Jeltsch, A., & Jurkowska, R. Z. (2015). *An ATPase-deficient variant of the SNF2 family member HELLS shows altered dynamics at pericentromeric heterochromatin. Journal of molecular biology*. http://doi.org/10.1016/j.jmb.2015.03.014

Lyko, F. (2017). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature Reviews Genetics*, *19*(2), 81–92. http://doi.org/10.1038/nrg.2017.80

MacLean, J. A., Chen, M. A., Wayne, C. M., Bruce, S. R., Rao, M., Meistrich, M. L., … Wilkinson, M. F. (2005). Rhox: A new homeobox gene cluster. *Cell*, *120*(3), 369–382. http://doi.org/10.1016/j.cell.2004.12.022

Maldonado-Saldivia, J., Bergen, J. Van Den, Krouskos, M., Gilchrist, M., Lee, C., Li, R., … Western, P. S. (2007). Dppa2 and Dppa4 Are Closely Linked SAP Motif Genes Restricted to Pluripotent Cells and the Germ Line. *Stem Cells*, (25), 19–28. http://doi.org/10.1634/stemcells.2006-0269.S

Malone, C. S., Miner, M. D., Doerr, J. R., Jackson, J. P., Jacobsen, S. E., Wall, R., & Teitell, M. (2001). CmC(A/T)GG DNA methylation in mature B cell lymphoma gene silencing. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(18), 10404–10409. http://doi.org/10.1073/pnas.181206898

Monk, M., Boubelik, M., & Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development*, *99*(3), 371–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3653008

Mozzetta, C., Boyarchuk, E., Pontis, J., & Ait-Si-Ali, S. (2015). Sound of silence: The properties and functions of repressive Lys methyltransferases. *Nature Reviews Molecular Cell Biology*, *16*(8), 499–513. http://doi.org/10.1038/nrm4029

Myant, K., & Stancheva, I. (2008). LSH cooperates with DNA methyltransferases to repress transcription. *Molecular and Cellular Biology*, *28*(1), 215–26. http://doi.org/10.1128/MCB.01073-07

Myant, K., Termanis, A., Sundaram, A. Y. M., Boe, T., Li, C., Merusi, C., … Stancheva, I. (2011). LSH and G9a / GLP complex are required for developmentally programmed DNA methylation. *Genome Research*, *21*, 83–94. http://doi.org/10.1101/gr.108498.110.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., … Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, *543*, 72–77. http://doi.org/10.1038/nature21373

Nishiyama, A., Yamaguchi, L., Sharif, J., Johmura, Y., Kawamura, T., Nakanishi, K., … Nakanishi, M. (2013). Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature*, *502*(7470), 249–253. http://doi.org/10.1038/nature12488

Niu, J., Chen, T., Han, L., Wang, P., Li, N., & Tong, T. (2011). Transcriptional activation of the senescence regulator Lsh by E2F1. *Mechanisms of Ageing and Development*, *132*(4), 180–6. http://doi.org/10.1016/j.mad.2011.03.004

Noh et al. (2015). Engineering of a histone-recognition domain in Dnmt3a alters the epigenetic landscape and phenotypic features of mouse ESCs. *Mol Cell*, *59*(1), 89–103. http://doi.org/10.1002/jmri.23741.Proton

Oda, M., Yamagiwa, A., Yamamoto, S., Nakayama, T., Tsumura, A., Sasaki, H., … Okano, M. (2006). DNA methylation regulates long-range gene silencing of an X-linked homeobox gene cluster in a lineage-specific manner. *Genes and Development*, *20*(24), 3382–3394. http://doi.org/10.1101/gad.1470906

Okano, M., Bell, D. W., Haber, D. a, & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, *99*(3), 247–57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10555141

Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., … Bestor, T. H. (2009). DNMT3L connects unmethylated lysine 4 of histone H3 to denovo methylation of DNA. *Nature*, *448*(7154), 714–717. http://doi.org/10.1038/nature05987.DNMT3L

Oppikofer, M., Bai, T., Gan, Y., Haley, B., Liu, P., Sandoval, W., … Cochran, A. G. (2017). Expansion of the ISWI chromatin remodeler family with new active complexes. *EMBO Reports*, *18*, 1697–1706. http://doi.org/10.15252/embr.201744011

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., … Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. *Current Biology*, *10*(8), 475–478. http://doi.org/10.1016/S0960-9822(00)00448-6

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., & Shirakawa, M. (2009). Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Reports*, *10*, 1235–1241. http://doi.org/10.1038/embor.2009.218

Raabe, E. H., Abdurrahman, L., Behbehani, G., & Arceci, R. J. (2001). An SNF2 Factor Involved in Mammalian Development and Cellular Proliferation. *Developmental Dynamics*, *221*, 92–105.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, *447*(7143), 425–32. http://doi.org/10.1038/nature05918

Reik, W., & Walter, J. (2001). GENOMIC IMPRINTING : PARENTAL INFLUENCE ON THE GENOME. *Nature Reviews Genetics*, *2*(21).

Ren, J., Briones, V., Barbour, S., Yu, W., Han, Y., Terashima, M., & Muegge, K. (2015). The ATP binding site of the chromatin remodeling homolog Lsh is required for nucleosome density and de novo DNA methylation at repeat sequences. *Nucleic Acids Research*, 1–12. http://doi.org/10.1093/nar/gku1371

Robertson, K. D., Uzvolgyi, E., Liang, G., Talmadge, C., Sumegi, J., Gonzales, F. A., & Jones, P. A. (1999). The human DNA methyltransferases (DNMTs) 1, 3a and 3b: Coordinate mRNA expression in normal tissues and overexpression in tumors. *Nucleic Acids Research*, *27*(11), 2291–2298. http://doi.org/10.1093/nar/27.11.2291

Rondelet, G., Dal Maso, T., Willems, L., & Wouters, J. (2016). Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. *Journal of Structural Biology*, 1–11. http://doi.org/10.1016/j.jsb.2016.03.013

Rountree, M. R., Bachman, K. E., & Baylin, S. B. (2000). DNMT1 binds HDAC2 and a new co-repressor , DMAP1 , to form a complex at replication foci. *Nature Genetics*, *25*, 269–277.

Rowe, H. M., & Trono, D. (2011). Dynamic control of endogenous retroviruses during development. *Virology*, *411*(2), 273–287. http://doi.org/10.1016/j.virol.2010.12.007

Russo V.E.A., Martienssen R.A., & Riggs A.D. (1996). Epigenetic Mechanisms of Gene Regulation. *Cold Spring Harbor Laboratory Press, Woodbury*.

Sarraf, S. A., & Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Molecular Cell*, *15*(4), 595–605. http://doi.org/10.1016/j.molcel.2004.06.043

Saxena, M., Stephens, M. A., Pathak, H., & Rangarajan, A. (2011). Transcription factors that mediate epithelial-mesenchymal transition lead to multidrug resistance by upregulating ABC transporters. *Cell Death and Disease*, *2*, e179. http://doi.org/10.1038/cddis.2011.61

Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences*, *103*(5), 1412–1417. http://doi.org/10.1073/pnas.0510310103

Schnütgen, F., De-Zolt, S., Van Sloun, P., Hollatz, M., Floss, T., Hansen, J., … von Melchner, H. (2005). Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(20), 7221–6. http://doi.org/10.1073/pnas.0502273102

Shaked, H., Avivi-Ragolsky, N., & Levy, A. A. (2006). Involvement of the arabidopsis SWI2/SNF2 chromatin remodeling gene family in DNA damage response and recombination. *Genetics*, *173*(2), 985–994. http://doi.org/10.1534/genetics.105.051664

Sharpless, N. E. (2006). Primary Murine Cells. *Cell Biology*, (13), 223–228.

Smallwood, S. A., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-pichon, A., … Andrews, S. R. (2012). Europe PMC Funders Group Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature Genetics*, *43*(8), 811–814. http://doi.org/10.1038/ng.864.Dynamic

Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H., Gnirke, A., Regev, A., & Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, *484*, 339–44. http://doi.org/10.1038/nature10960

Smith, Z. D., & Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nature Reviews. Genetics*, *14*(3), 204–20. http://doi.org/10.1038/nrg3354

Sormani, G., Haerter, J. O., Lövkvist, C., & Sneppen, K. (2016). Stabilization of epigenetic states of CpG islands by local cooperation. *Mol. BioSyst.* http://doi.org/10.1039/C6MB00044D

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., … Schübeler, D. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, *480*(7378), 490–495. http://doi.org/10.1038/nature10716

Stockdale, C., Flaus, A., Ferreira, H., & Owen-hughes, T. (2006). ANALYSIS OF NUCLEOSOME REPOSITIONING BY YEAST ISWI AND CHD1 CHROMATIN REMODELLING COMPLEXES. *Journal of Biological Chemistry*, *281*(24), 16279–16288. http://doi.org/10.1074/jbc.M600682200.ANALYSIS

Stützer, A., Liokatis, S., Kiesel, A., Schwarzer, D., Sprangers, R., Söding, J., … Fischle, W. (2016). Modulations of DNA Contacts by Linker Histones and Post-translational Modifications Determine the Mobility and Modifiability of Nucleosomal H3 Tails. *Molecular Cell*, *61*(2), 247–259. http://doi.org/10.1016/j.molcel.2015.12.015

Sun, L.-Q., Lee, D. W., Zhang, Q., Xiao, W., Raabe, E. H., Meeker, A., … Arceci, R. J. (2004). Growth retardation and premature aging phenotypes in mice with disruption of the SNF2-like gene, PASG. *Genes & Development*, *18*(9), 1035–46. http://doi.org/10.1101/gad.1176104

Takeshima, H., Suetake, I., Shimahara, H., Ura, K., Tate, S. I., & Tajima, S. (2006). Distinct DNA methylation activity of Dnmt3a and Dnmt3b towards naked and nucleosomal DNA. *Journal of Biochemistry*, *139*(3), 503–515. http://doi.org/10.1093/jb/mvj044

Tao, Y., Xi, S., Shan, J., Maunakea, A., Che, A., Briones, V., … Robert, M. (2011). Correction for Tao et al., Lsh, chromatin remodeling family member, modulates genome-wide cytosine methylation patterns at nonrepeat sequences. *Proceedings of the National Academy of Sciences*, *108*(37), 15535–15535. http://doi.org/10.1073/pnas.1112647108

Termanis, A., Torrea, N., Culley, J., Kerr, A., Ramsahoye, B., & Stancheva, I. (2016). The SNF2 family ATPase LSH promotes cell-autonomous *de novo* DNA methylation in somatic cells. *Nucleic Acids Research*, gkw424. http://doi.org/10.1093/nar/gkw424

Thijssen, P. E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., … Sasaki, H. (2015). Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. *Nature Communications*, *6*, 7870. http://doi.org/10.1038/ncomms8870

Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., … Bird, A. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*, *464*(7291), 1082–1086. http://doi.org/10.1038/nature08924

Tremethick, D. J. (2007). Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. *Cell*, *128*(4), 651–654. http://doi.org/10.1016/j.cell.2007.02.008

Trojer, P., & Reinberg, D. (2007). Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Molecular Cell*, *28*(1), 1–13. http://doi.org/10.1016/j.molcel.2007.09.011

Uysal, F., Akkoyunlu, G., & Ozturk, S. (2015). Dynamic expression of DNA methyltransferases (DNMTs) in oocytes and early embryos. *Biochimie*, *116*, 103–113. http://doi.org/10.1016/j.biochi.2015.06.019

van Kruijsbergen, I., Hontelez, S., & Veenstra, G. J. C. (2015). Recruiting Polycomb to chromatin. *Int J Biochem Cell Biol*, *67*, 177–187. http://doi.org/10.1080/10937404.2015.1051611.INHALATION

Vann, K. R., & Kutateladze, T. G. (2017). Histone H3 Dual Ubiquitylation Mediates Maintenance DNA Methylation. *Molecular Cell*, *68*, 261–262. http://doi.org/10.1016/j.molcel.2017.10.007

von Eyss, B., Maaskola, J., Memczak, S., Möllmann, K., Schuetz, A., Loddenkemper, C., … Ziebold, U. (2012). The SNF2-like helicase HELLS mediates E2F3-dependent transcription and cellular transformation. *The EMBO Journal*, *31*(4), 972–85. http://doi.org/10.1038/emboj.2011.451

Vongs, A., Kakutani, T., Martienssen, R. a, & Richards, E. J. (1993). Arabidopsis thaliana DNA methylation mutants. *Science*, *260*(5116), 1926–1928. http://doi.org/10.1126/science.8316832

Waddington, C. H. (1957). The Strategy of the Genes. *(Allen and Unwin, London)*.

Wade, P. A., & Wolffe, A. P. (2001). ReCoGnizing methylated DNA. *Nature Structural Biology*, *8*(7), 575–577. http://doi.org/10.1038/89593

Walter, M., Teissandier, A., Pérez-Palacios, R., & Bourc'his, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife*, 1–30. http://doi.org/10.1017/CBO9781107415324.004

Wang, R., Shi, Y., Chen, L., Jiang, Y., Mao, C., Yan, B., … Wang, X. (2015). The ratio of FoxA1 to FoxA2 in lung adenocarcinoma is regulated by LncRNA HOTAIR and chromatin remodeling factor LSH. *Scientific Reports*, *5*, 1–11. http://doi.org/10.1038/srep17826

Wang, W., Yang, J., Liu, H., Lu, D., Chen, X., Zenonos, Z., … Liu, P. (2011). Rapid and efficient reprogramming of somatic cells to induced pluripotent stem cells by retinoic acid receptor gamma and liver receptor homolog 1. *Proceedings of the National Academy of Sciences*, *108*(45), 18283–18288. http://doi.org/10.1073/pnas.1100893108

Wang, Z., Zang, C., Rosenfeld, J. a, Schones, D. E., Cuddapah, S., Cui, K., … Michael, Q. (2009). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics*, *40*(7), 897–903. http://doi.org/10.1038/ng.154.Combinatorial

Waseem, A., Ali, M., Odell, E. W., Fortune, F., & Teh, M. T. (2010). Downstream targets of FOXM1: CEP55 and HELLS are cancer progression markers of head and neck squamous cell carcinoma. *Oral Oncology*, *46*, 536–542. http://doi.org/10.1016/j.oraloncology.2010.03.022

Weber, C. M., & Henikoff, S. (2014). Histone variants:dynamic punctuation in transcription. *Genes & Development*, (28), 672–682. http://doi.org/10.1101/gad.238873.114.Freely

Wen, H., Li, Y., Xi, Y., Jiang, S., Stratton, S., Peng, D., … Shi, X. (2014). ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. *Nature*, *508*, 263–8. http://doi.org/10.1038/nature13045

Whitehouse, I., Rando, O. J., Delrow, J., & Tsukiyama, T. (2007). Chromatin remodelling at promoters suppresses antisense transcription. *Nature*, *450*(7172), 1031–5. http://doi.org/10.1038/nature06391

Woodcock, C. L. (1994). Chromatin fibers observed in situ in frozen hydrated sections. Native fiber diameter is not correlated with nucleosome repeat length. *Journal of Cell Biology*, *125*(1), 11–19. http://doi.org/10.1083/jcb.125.1.11

Woodcock, C. L., & Dimitrov, S. (2001). Higher-order structure of chromatin and chromosomes. *Current Opinion in Genetics & Development*, *11*, 130–135.

Woodcock, C. L. F., Safer, J. P., & Stanchfield, J. E. (1976). Structural repeating units in chromatin. I. Evidence for their general occurrence. *Experimental Cell Research*, *97*(1), 101–110. http://doi.org/10.1016/0014-4827(76)90659-5

Wu, H., Thijssen, P. E., de Klerk, E., Vonk, K. K. D., Wang, J., den Hamer, B., … Daxinger, L. (2016). Converging disease genes in ICF syndrome: ZBTB24 controls expression of CDCA7 in mammals. *Human Molecular Genetics*, *25*(18), 4041–4051. http://doi.org/10.1093/hmg/ddw243

Xi, S., Geiman, T. M., Briones, V., Tao, Y. G., Xu, H., & Muegge, K. (2009). Lsh Participates in DNA Methylation and Silencing of Stem Cell Genes. *Stem Cells*, (27), 2691–2702. http://doi.org/10.1002/July

Xi, S., Zhu, H., Xu, H., Schmidtmann, A., Geiman, T. M., & Muegge, K. (2007). Lsh controls Hox gene silencing during development. *Proc Natl Acad Sci U S A*, *104*(36), 14366–14371. http://doi.org/0703669104 [pii]\r10.1073/pnas.0703669104

Xia, C. Q., & Smith, P. G. (2012). Drug efflux transporters and multidrug resistance in acute leukemia: therapeutic impact and novel approaches to mediation. *Molecular Pharmacology*, *82*(6), 1008–21. http://doi.org/10.1124/mol.112.079129

Xu, G.-L., Bestor, T. H., Bourc'his, D., Hsieh, C.-L., Tommerup, N., Bugge, M., … Viegas-Péquignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature*, *402*(6758), 187–191. http://doi.org/10.1038/46052

Xu, M., & Zhu, B. (2010). Nucleosome assembly and epigenetic inheritance. *Protein & Cell*, *1*(9), 820–829. http://doi.org/10.1007/s13238-010-0104-0

Yang, J., Wang, W., Ooi, J., Campos, L. S., Lu, L., & Liu, P. (2015). Signalling Through Retinoic Acid Receptors is Required for Reprogramming of Both Mouse Embryonic Fibroblast Cells and Epiblast Stem Cells to Induced Pluripotent Stem Cells. *Stem Cells (Dayton, Ohio)*, *33*(5), 1390–404. http://doi.org/10.1002/stem.1926

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., … Taipale, J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, *356*(6337). Retrieved from http://science.sciencemag.org/content/356/6337/eaaj2239.abstract

Yoder, J. A., Soman, N. S., Verdine, G. L., & Bestor, T. H. (1997). DNA ( cytosine-5 )-methyltransferases in Mouse Cells and Tissues . Studies with a Mechanism-based Probe. *Journal of Molecular Biology*, *270*, 385–395.

Yoder, J. A., Walsh, C. P., & Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics*, *13*(8), 335–340. http://doi.org/10.1016/S0168-9525(97)01181-5

Yu, W., McIntosh, C., Lister, R., Zhu, I., Han, Y., Ren, J., … Muegge, K. (2014). Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. *Genome Research*. http://doi.org/10.1101/gr.172015.114

Zhang, P., Su, L., Wang, Z., Zhang, S., Guan, J., Chen, Y., … Li, Z. (2012). The Involvement of 5-Hydroxymethylcytosine in Active DNA Demethylation in Mice1. *Biology of Reproduction*, *86*(4), 1–9. http://doi.org/10.1095/biolreprod.111.096073

Zhang, T., Termanis, A., Özkan, B., Bao, X. X., Culley, J., de Lima Alves, F., … Stancheva, I. (2016). G9a/GLP Complex Maintains Imprinted DNA Methylation in Embryonic Stem Cells. *Cell Reports*, 77–85. http://doi.org/10.1016/j.celrep.2016.03.007

Zhou, R., Han, L., Li, G., & Tong, T. (2009). Senescence delay and repression of p16INK4a by Lsh via recruitment of histone deacetylases in human diploid fibroblasts. *Nucleic Acids Research*, *37*(15), 5183–96. http://doi.org/10.1093/nar/gkp533

Zhu, H., Geiman, T. M., Xi, S., Jiang, Q., Schmidtmann, A., Chen, T., … Muegge, K. (2006). Lsh is involved in de novo methylation of DNA. *The EMBO Journal*, *25*(2), 335–45. http://doi.org/10.1038/sj.emboj.7600925